ENZYMATIC HYDROLYSIS OF ALKALINE PRETREATED BIOMASSES: ASSESSMENT OF COMMERCIAL HEMICELLULASE MIXTURES AND THE USE OF FTIR SPECTROSCOPY TO PREDICT SACCHARIFICATION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Deborah Lynne Sills

August 2011

ENZYMATIC HYDROLYSIS OF ALKALINE PRETREATED BIOMASSES: ASSESSMENT OF COMMERCIAL HEMICELLULASE MIXTURES AND THE USE OF FTIR SPECTROSCOPY TO PREDICT SACCHARIFICATION

Deborah Lynne Sills, Ph. D.

Cornell University 2011

Three studies that address plant cell wall recalcitrance and its interaction with bioconversion processes are presented in this dissertation. First, the effects of commercial cellulase and hemicellulase mixtures on sugar productions from two perennial biomasses were measured. A low-impact, high-diversity mixture of prairie biomasses (LIHD) and switchgrass were each subjected to NaOH pretreatment, followed by hydrolysis with a commercial cellulase and β-glucosidase mixture supplemented with either of two hemicellulases. A novel modeling approach was presented that used "marginal effectiveness" (incremental grams sugar per incremental mg enzyme) to determine enzyme loadings required for optimal sugar productions. Results suggested that there was no need to customize cellulase loading or hemicellulase supplementation for NaOH-pretreated switchgrass and LIHD.

Second, Fourier transform infrared spectroscopy (FTIR) combined with partial least squares (PLS) regression was used to estimate glucose and xylose productions from NaOH pretreatment and enzymatic hydrolysis of six plant biomasses that represent a variety of potential biofuel feedstocks. Two switchgrass cultivars, big bluestem grass, LIHD, mixed hardwood, and corn

stover were subjected to four levels of NaOH pretreatment to produce 24 samples with a wide range of potential digestibility. PLS models were constructed by correlating FTIR spectra of pretreated samples to measured values of glucose and xylose conversions (g sugar per 100 g potential sugar) and yields (g sugar per 100 g TS).

Third, PLS models constructed from FTIR spectra of the six raw biomasses plus four levels of pretreatment (0, 5, 10, and 20 g NaOH per 100 g TS) satisfactorily predicted solubilizations of plant cell wall constituents through pretreatment. Additionally, PLS models constructed from FTIR spectra of the six raw biomasses plus three levels of pretreatment (5, 10, and 20 g NaOH per 100 g TS) accurately predicted overall glucose and xylose conversions (g sugar per 100 g potential sugar) and yields (g sugar per 100 g TS) from pretreatment plus enzymatic hydrolysis. The ability to predict sugar yields without prior knowledge of biomass composition suggests that FTIR combined with PLS regression may be able to replace wet chemical analyses and enzymatic assays in estimating saccharification from lignocellulosic biomass.

BIOGRAPHICAL SKETCH

Before attending college, Deborah served for two years in the Israeli military, traveled extensively in Southeast Asia, and gave birth to her daughter, Noa. Deborah received an undergraduate degree from Montana State University in Civil Engineering with an emphasis on Bio-resources in 2001, after which she came to Cornell University to attend graduate school. Deborah researched aerobic biodegradation of organic pollutants under the direction of James Gossett and received a Masters Degree in 2004. Deborah continued on to pursue a Ph.D, focusing her research on bioconversion of lignocellulosic biomass. During her graduate studies at Cornell, Deborah especially enjoyed teaching undergraduate students and riding her bicycle up and down the hills of Ithaca.

ACKNOWLEDGMENTS

First and foremost I wish to thank my major academic advisor, James Gossett, for challenging, guiding, and supporting me. I cannot imagine a better mentor. I am also grateful to my committee members, Larry Walker and David Wilson. Larry generously welcomed me into his lab, provided me with analytical equipment and supplies, and taught me how to run a clean and orderly laboratory. David Wilson was always willing to share his immense knowledge on cellulases and insightful thoughts and advice on scientific research.

I thank Diana Irwin from the Wilson Lab for instruction on biochemical assays, James Robertson and Deborah Ross from the Department of Animal Science for instruction on compositional analysis of lignocellulosic biomass, Stephane Corgié for his assistance with FTIR analysis, Karl Siebert for assistance with PLS regression, and Bryan Hanson from DePauw University for teaching me how to use R. I am grateful to my friends and fellow students for their help and moral support: Hnin Aung, Jake Bolewski, Ed Evans, Ben Heavner, Gretchen Heavener, Brian King, Po-Hsun Lin, Jeremy Luterbacher, and Sarah Munro.

Finally, I thank my mother, Leslie Banks-Sills, for her invaluable and tireless assistance in completing this dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Context

Increasing energy demands, diminishing fossil fuel resources, and the need to curb greenhouse gas emissions have sparked renewed interest in developing a biomass-based economy. Like petroleum refineries, which convert crude oil into a multitude of fuels and chemicals, biorefineries are expected to transform plant-biomass feedstocks into fuels, power, and chemicals (Dodds and Gross 2007; Faust et al. 2008; Kamm and Kamm 2004). Plant cell walls, constructed from lignocellulose, represent the largest source of renewable organic carbon on earth, and most of it is untapped (Pauly and Keegstra 2008). Furthermore, a study conducted by the United States Department of Energy reported that that about 1.3 billion tons of non-edible plant biomass can be produced each year in addition to current agricultural and forestry productions (Somerville 2006). If all of this biomass were converted into fuel ethanol — as an example — the United States could satisfy 30% of its demand for transportation fuel (Somerville 2006).

Biorefineries are expected to work with a wide variety of feedstocks. These include agricultural residues such as corn stover and wheat straw as well as dedicated energy crops such as perennial grasses (e.g. switchgrass), and soft and hard woods. Additionally, there is interest in feedstocks comprised of multiple plant types grown together. Tilman et al. (2006) reported that a low-impact, high-diversity mixture of perennial prairie biomasses (LIHD) exhibited higher yields growing on marginal land without agro-chemical inputs compared to switchgrass. This

suggests that cultivating mixed species may increase overall sustainability of biorefinery products.

The conversion of lignocellulosic feedstocks into simple sugars lies at the center of the biorefinery. One promising technology, referred to as biochemical conversion, depolymerizes plant cell walls into simple sugars via two processes — pretreatment followed by cellulase- and hemicellulase-mediated hydrolysis. Biochemical saccharification of plant cell walls is the biggest technical bottleneck preventing the development of full-scale biorefineries (Jørgensen et al. 2007).

Enzymatic deconstruction of carbohydrate-rich, pretreated plant biomasses is limited by the recalcitrance of plant cell walls. While it is known that the structure of cellulose, the presence of hemicellulose and lignin, and cross-links that hold cell-wall polymers together contribute to recalcitrance, exact chemical and physical mechanisms by which lignocellulose resists hydrolysis have not been elucidated. Traditional wet-chemistry techniques used to measure the chemical make-up of lignocellulose quantify three major polymer types in lignocellulose — cellulose, hemicellulose, and lignin. However, such analyses do not reveal subtle details associated with the chemical make-up and structure that influence recalcitrance, and in turn, sugar productions from combined pretreatment and enzymatic hydrolysis.

1.2 Objectives

It is important to enhance our understanding of plant cell wall recalcitrance and its interaction with saccharification processes. New approaches are needed for the design and optimization of pretreatment and hydrolysis conditions. Additionally, limitations of wet

chemistry methods need to be addressed in order to increase capabilities of predicting sugar yields from biochemical conversion processes.

Two main objectives are addressed in this dissertation. The first objective was to address the following question: Must hydrolysis be customized to two biomass substrates that have the potential to be major feedstocks for biorefineries — switchgrass and LIHD (Tilman et al. 2006)? We addressed this question through a study that measured the effect of adding two commercial hemicellulase mixtures to cellulase-mediated hydrolysis of NaOH pretreated switchgrass and LIHD. In this study, we also addressed the concept of optimal sugar production, and presented a novel modeling approach and the use of "marginal effectiveness" as a way of looking at enzyme loadings.

The second objective was to ascertain the potential of using Fourier transform infrared spectroscopy (FTIR) combined with partial least squares (PLS) regression to predict effects of NaOH pretreatment on lignocellulosic biomass and subsequent saccharification. Two studies were conducted with six biomass types that represent a variety of potential plant feedstocks — two switchgrass cultivars, big bluestem grass, LIHD, mixed hardwood, and corn stover. The first of the FTIR studies predicted saccharification through enzymatic hydrolysis using 24 FTIR spectra of NaOH-pretreated biomasses. The second study predicted solubilizations of plant cell wall constituents through pretreatment as well as overall sugar productions from combined pretreatment and enzymatic hydrolysis from spectra of the six raw biomasses plus level of pretreatment.

1.3 Organization

In Chapter 2, a literature review is presented on topics related to bioconversion of lignocellulosic biomass, FTIR spectroscopy, and PLS regression. In Chapter 3, a study is presented that measured the effect of adding two commercial hemicellulase mixtures to cellulase-mediated hydrolysis of NaOH pretreated switchgrass and LIHD. In Chapters 4 and 5, two studies are presented that use FTIR to construct PLS regression models that estimate sugars released in pretreatment and enzymatic hydrolysis. Finally, a summary and conclusions that outline the significance of this work, as well as suggestions for future studies, are presented in Chapter 6.

REFERENCES

Dodds DR, Gross RA. 2007. Chemistry: Chemicals from biomass. Science 318:1250-1251.

Faust TD, Ibsen KN, Dayton DC, Hess JR, Kenney KE. 2008. The Biorefinery. In: Himmel ME, editor. Biomass recalcitrance. Deconstructing the plant cell wall for bioenergy. Oxford, United Kingdom: Blackwell Publishing Ltd. p 7-37.

Jørgensen H, Kristensen JB, Felby C. 2007. Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. Biofuels, Bioprod Bior 1:119-134.

Kamm B, Kamm M. 2004. Principles of biorefineries. Appl Microbiol Biotechnol 64:137-145.

Pauly M, Keegstra K. 2008. Cell-wall carbohydrates and their modification as a resource for biofuels. Plant J 54:559-568.

Somerville C. 2006. The billion-ton biofuels vision. Science 312:1277.

Tilman D, Hill J, Lehman C. 2006. Carbon negative biofuels from low-input high-diversity grassland biomass. Science 314:1598-1600.

CHAPTER 2

BACKGROUND

2.1 Plant Cell Walls — Lignocellulose

Plant cell walls are composed of lignocellulose — a composite material that provides structural integrity to plants, as well as resistance to microbial degradation. Lignocellulose is composed of 30-50% cellulose (a glucose polymer), 15-35% hemicellulose (non-cellulosic carbohydrate polymers), and 10-30% lignin (phenylpropanoid polymers) (Pauly and Keegstra 2008). Additional minor components of cell walls include soluble sugars (1-5%), protein (3-10%), lipids (1-5%), and ash (5-10%) (Pauly and Keegstra 2008).

2.1.1 Cellulose

Cellulose consists of long chains of 400 to 14,000 D-glucose monomers linked with β-1,4-glycosidic bonds. Each chain has a reducing end, consisting of an open hemiacetal ring with an exposed reducing aldehyde, and a non-reducing end with a closed hemiacetyl ring (see Fig. 2.1). Furthermore, each glucose monomer in cellulose is rotated 180º with respect to its neighbor, and the chain is made up of repeating units of the disaccharide, cellobiose (see Fig. 2.1).

A cellulose microfibril (3–5 nm in thickness) typically is made up of 36 cellulose chains, that are formed simultaneously in cell walls (Mohnen et al. 2008). As the microfibril is polymerized by cellulose synthase (e.g. CesA), cellulose chains twist into long-period, helical structures (Atalla et al. 2008). Adjacent chains are held together through inter- and intra-chain

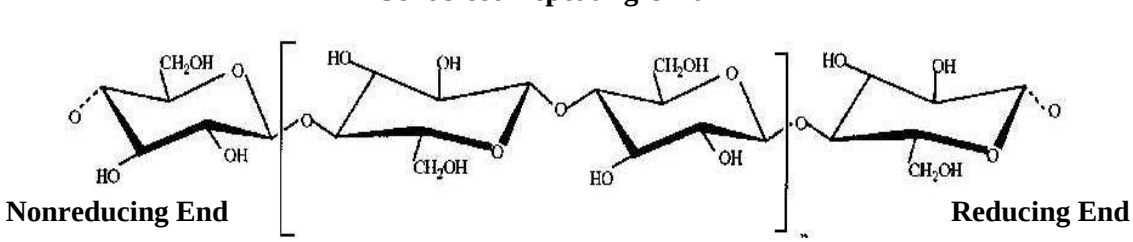

Figure 2.1 Cellulose chain composed of ß-1,4-linked D-glucose units. Adapted from Brown Jr. et al. (1996).

hydrogen bonds, as well as van der Waals forces, resulting in hydrophobic crystalline microfibrils that are resistant to chemical and enzymatic hydrolysis. The abundance and mobility of cellulose-polymerizing enzymes determine the orientation, and thickness of fibers in different types of cells. Cellulose in plant cell walls exists in crystalline, semi-crystalline, and amorphous forms. Crystalline cellulose in plants is composed primarily of the polymorph Cellulose $I_ß$, with small amounts of the I_α polymorph (Atalla et al. 2008; Kuga and Brown 1991). Cellulose $I_ß$ is considered to be the polymorph most resistant to hydrolysis (Qiang et al. 2005). It is more tightly packed than other forms of cellulose, such as Cellulose II, due to differences in hydrogen-bonding patterns (O'Sullivan 1997). Cellulose I can be converted to Cellulose II through treatment with strong alkali (Nelson and O'Connor 1964).

2.1.2 Hemicellulose

Hemicellulose is a term that defines non-cellulosic carbohydrate polymers that can be extracted from plant cell walls with alkali (Mohnen et al. 2008). In contrast to cellulose, hemicellulose composition varies among plant species, organs (leaf and stem), and tissues

[xylem, sclerenchyma, parenchyma, and epidermis] (Ebringerova 2006). Hemicellulose polymers are composed of pentoses (xylose and arabinose), hexoses (glucose, mannose, and galactose), uronic acids (4-O-methyl glucuronic and galacturonic acids), and low-molecular weight *p*-hydroxycinnamic acids (ferulic and coumaric acids). Hemicellulose chains are typically short chains, and they can be linear or highly branched with multiple side groups.

There are four main polymer types that make up most hemicelluloses: xylan, xyloglucan, mixed-linkage glucans, and mannans including glucomannans and galactomannans (Harris and Stone 2008). Xylan consists of a ß-1,4-xylan backbone, with or without glucuronic-acid and arabinose side chains. Xyloglucan has a ß-1,4-glucan backbone that is highly substituted with xylose side chains that may also contain galactose, arabinose, or fucose. Mixed-linkage glucan is made up of D-glucose monomers linked via ß-1,3 and ß-1,4 bonds, and mannan consists of a ß-1,4-mannan backbone with α-1,6 galactose branches — or, in the case of glucomannan, a backbone with both glucose and mannose residues linked with ß-1,4 linkages.

Xylan is the predominant hemicellulose polymer in perennial grasses, corn stover, and wood (Gilbert et al. 2008). Arabinoxylan, the most common xylan polymer in grasses, is presented in Fig. 2.2 as an exemplary hemicellulose polymer. Arabinoxylan consists of a β-1,4-D-xylose backbone and α -L-arabinose subunits usually linked to the O-2 and the O-3 position of the xylose residues. Xylose residues may also be modified by acetylation in the O-2 or O-3 positions. In addition, ferulic and *p*-coumaric acid may be linked via ester bonds to the O-5 position of arabinose side chains (Carpita and McCann 2000; Gilbert et al. 2008). Arabinoxylan plays a major role in maintaining the structural integrity of the plant cell walls in grass species (Akin 2007; Gilbert et al. 2008).

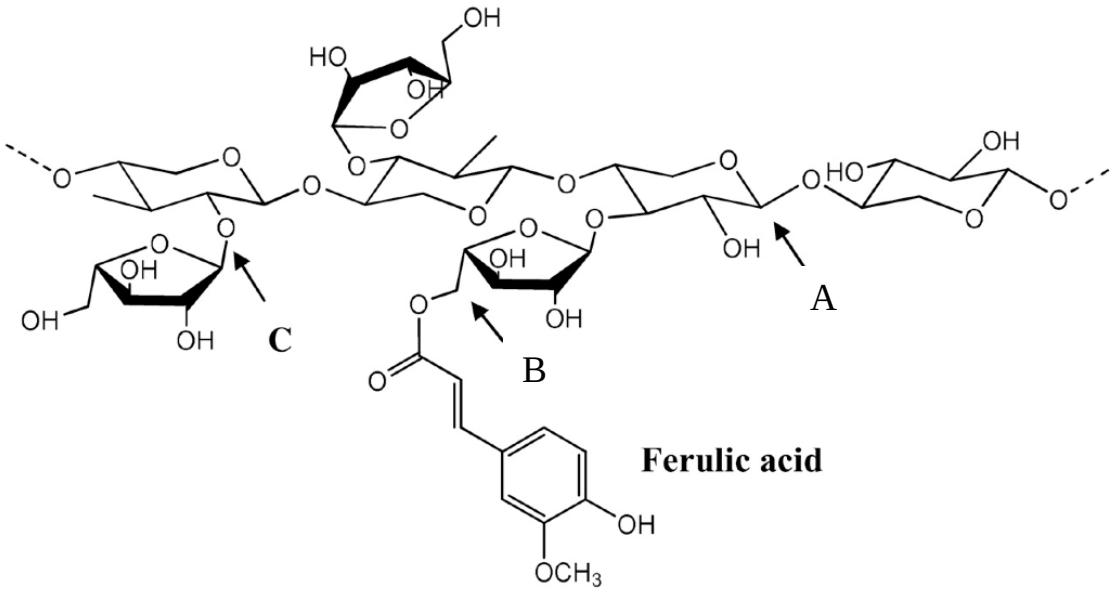

Figure 2.2 Structure of arabinoxylan. Arabinoxylan consists of a ß-1,4-linked xylan backbone (A); with α-1,2-linked arabinose side chains (B); and ferulic acid linked to the O-5 position of arabinose side-chains (C). Adapted from de O. Buanafina (2009).

2.1.3 Lignin

Lignin, the second (after cellulose) most abundant organic compound in nature, is a three-dimensional, amorphous, and complex phenolic compound. Lignin strengthens cell walls, facilitates water transport within plants, and coats and protects cell-wall polysaccharides from degradation (Davin et al. 2008).

Lignin is derived from three phenylpropane monomers, referred to as monolignols — *p*-coumaryl, coniferyl, and sinapyl alcohols [see Figs. 2.3 (a), (b), and (c)] — which are precursors to three types of lignin polymers — *p*-hydroxyphenyl (H), guaicyl (G), and syringyl (S) lignins [see Figs. 2.3 (d), (e), and (f)]. As in the case of hemicellulose, ratios and absolute

amounts of lignin monomers vary widely among plant species, organs, and tissues (Besle et al. 1994; Davin et al. 2008). Furthermore, in grasses, lignin may also contain smaller amounts (up to 10% of total lignin) of *p*-hydroxycinnamate esters (feruloyl and coumaryl esters) [see Figs. 2.3 (g), (h), (i)] (Ralph 2010). There is disagreement in the literature whether hydroxycinnamate esters are true lignin precursors or not, and these molecules are often included in descriptions of hemicellulose or referred to as "non-core" lignin.

Lignin is polymerized in the cell wall via oxidative (radical mediated) coupling reactions dictated primarily by the chemistry of the three monolignols [see Figs. 2.3 (a), (b), and (c)], as well as the chemistry of intermediate oligomers (Hatfield and Vermerris 2001). Common types of coupling reactions that polymerize lignin form ß-O-4 (also referred to as ß-ether linkages), ß-5, ß-ß, 5-5, and 5-O-4 linkages (Ralph 2010). Two dimers, each comprised of two coniferyl-alcohol monomers formed with 5-5- and 5-O-4- coupling reactions are presented in Fig. 2.4, as examples of lignin-lignin bonds.

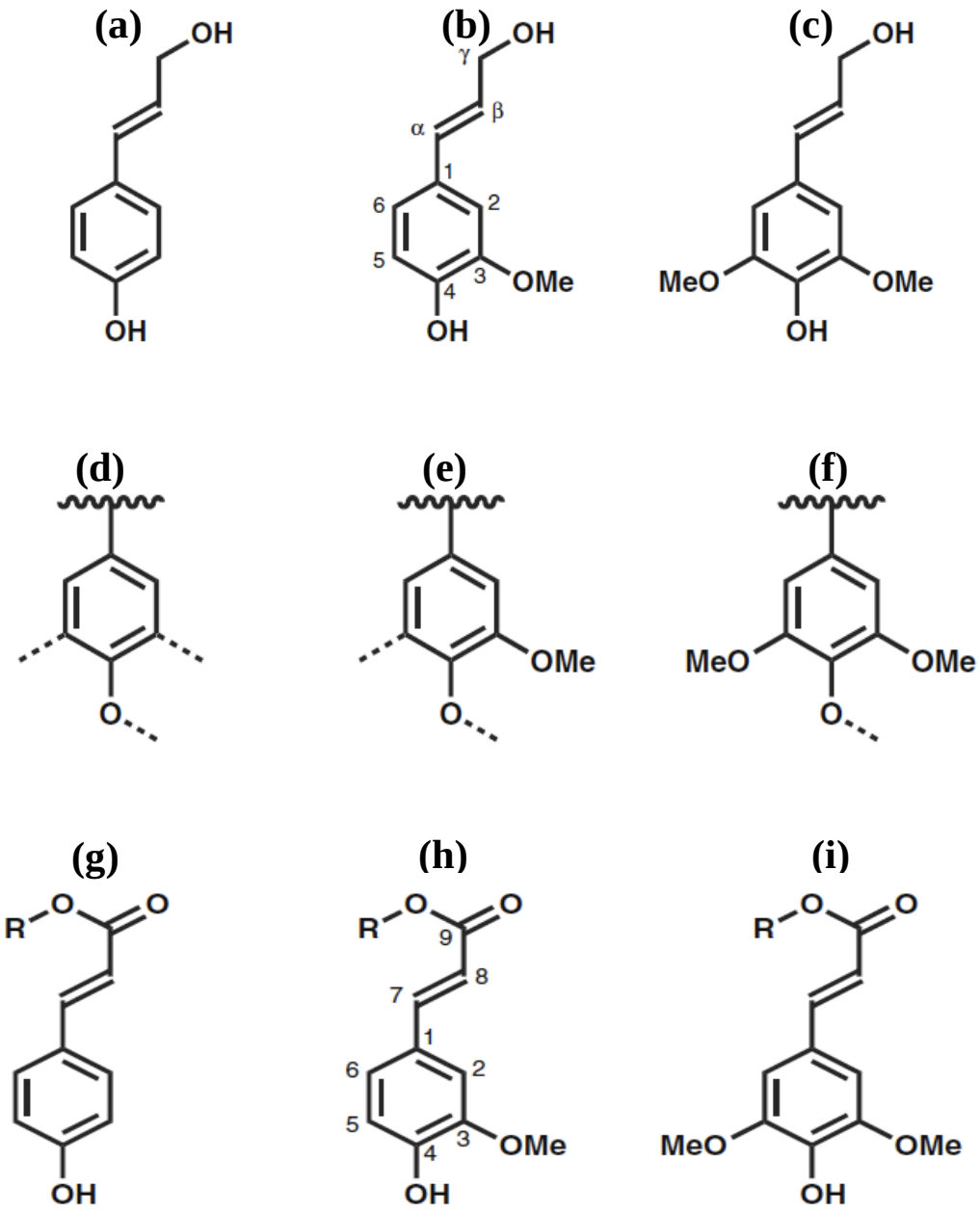

Figure 2.3 Primary lignin monomers: (a) *p*-coumaryl, (b) coniferyl, and (c) sinapyl alcohols; primary lignin subunits: (d) *p*-hydroxyphenyl (H), (e) guaicyl (G), and (f) syringyl (S) lignin; and hydroxycinnamate esters: (g) *p*-coumarate, (h) ferulate, and (i) sinapate. Adapted from Ralph (2010).

(a) 5-5-dimer **(b) 5-O-4-dimer**

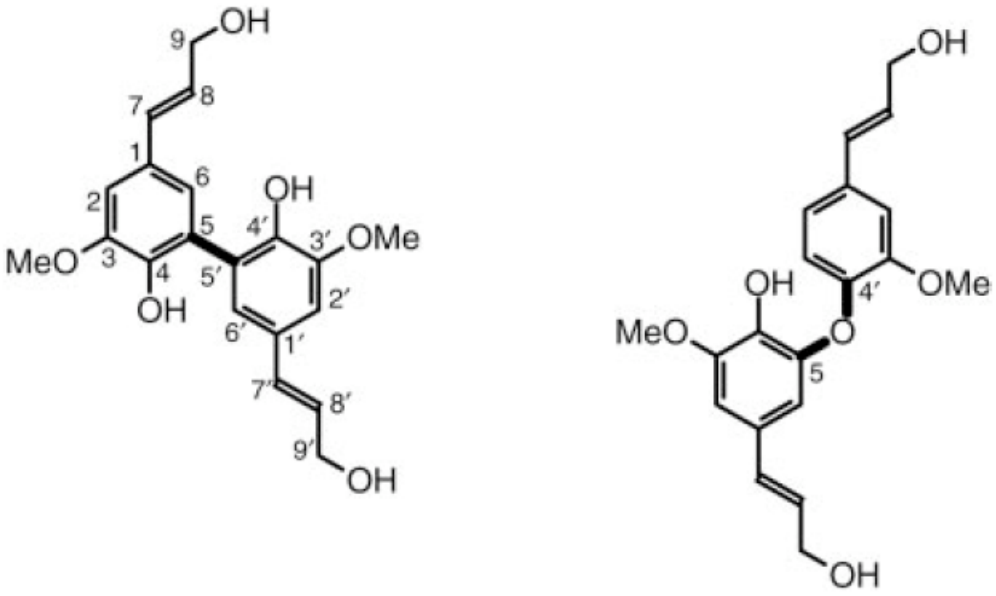

Figure 2.4 Lignin dimers composed from two coniferyl-alcohol monomers via (a) 5-5- and (b) 5-O-4- coupling reactions. Bonds that link two monolignols are in bold, and monolignol carbon numbers are shown in (a) to illustrate the conventional numbering scheme. Only the carbons participating in the bond are labeled in (b). Adapted from Hatfield and Vermerris (2001).

2.1.4 Physical and Chemical Structure of Plant Cell Walls

There are two main types of plant cell walls — primary and secondary. Primary walls are formed while the plant cell is growing and they consist of cellulose microfibrils embedded in a gel matrix composed of non-cellulosic carbohydrates (Harris and Stone 2008). Additionally, primary-cell-wall proteins (e.g. extensins and transglycosylases) provide activities necessary for the construction of elongating cell walls.

Secondary, lignified, walls form in mature plant cells, and they make up the majority of lignocellulosic biomass available for bioprocessing. A simplistic model of a secondary grass cell

wall is presented in Fig. 2.5. In secondary cell walls, densely packed and ordered cellulose microfibrils are enmeshed in a polysaccharide-lignin matrix. Hemicellulose polymers are often sandwiched between cellulose microfibrils, and hydrogen bonds result in tight associations between hemicellulose and cellulose (Atalla et al. 1993). Additionally, lignin is covalently bound to hemicellulose through lignin-carbohydrate complexes [LCC] (Ralph 2010), consisting of ester and ether bonds. LCC bonds are known to limit biological degradation of forage crops (Casler and Jung 2006). The composite structure of lignocellulose provides plants with cell walls that are extremely resistant to chemical and biological degradation. The development of full-scale biorefineries is dependent on overcoming the recalcitrance of plant cell walls.

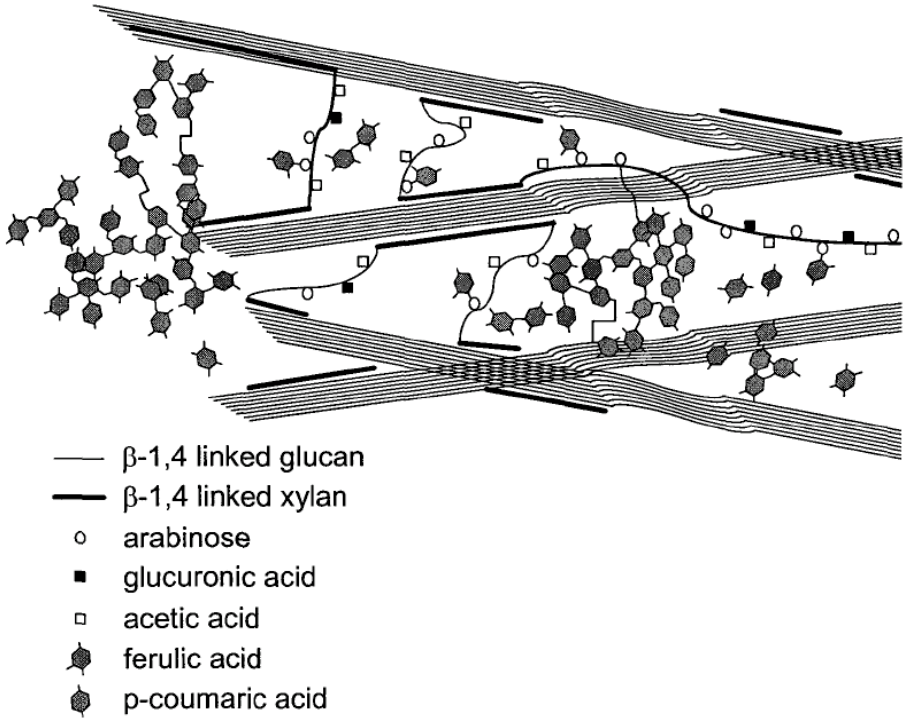

Figure 2.5 Model of a grass cell wall. Crystalline (straight parallel lines) and amorphous (wavy curved lines) cellulose is enmeshed in a hemicellulose-lignin matrix. Unbranched xylan polymers (short thick lines) are hydrogen-bonded to cellulose, while branched xylan polymers with side groups (arabinosyl, glucuronyl and acetyl groups are depicted with circles and squares) cross-link cellulose microfibrils. Lignin-carbohydrate bonds that link arabinose to ferulic and coumaric acids further strengthen the cell wall matrix. Lignin consists of cross-linked lignol monomers — hydroxyphenyl, guaicyl, and syringyl units (Chundawat 2009).

2.2 Biochemical Deconstruction of Plant Cell Walls

Plants have evolved for thousands of years and developed cell walls that resist microbial degradation. Both physical and chemical features of lignocellulose limit accessibility of cellulolytic enzymes to cellulose and must be overcome through pretreatment. Physical impediments include inaccessible pores (Grethlein 1985), low accessible surface area (Grethlein 1985) and cellulose crystallinity (Fan et al. 1981). Chemical limitations include the presence of

hemicellulose (Ohgren et al. 2007), lignin (Grabber 2005), and lignin-carbohydrate linkages (Casler and Jung 2006; Grabber 2005). Lignin reduces rates and ultimate sugar yields in enzymatic hydrolysis of pretreated biomass by two primary mechanisms: physical blocking of carbohydrate polymers (Mansfield et al. 1999), and adsorbing and inactivating cellulases (Berlin et al. 2006; Eriksson et al. 2002; Ooshima et al. 1990; Palonen 2004; Sewalt et al. 1997).

Depolymerizing plant cell walls into simple sugars is thus a formidable challenge for engineers and scientists. Currently, the most promising strategy is a biochemical platform that includes two major processes: thermochemical pretreatment followed by cellulase- and hemicellulase-mediated hydrolysis (Faust et al. 2008; Kamm and Kamm 2004).

2.2.1 Pretreatment

Pretreatment is required to achieve high saccharification yields in subsequent enzymatic hydrolysis. The main goal of pretreatment is to overcome plant cell wall recalcitrance by opening up the lignocellulosic matrix and increasing the accessibility of cell-wall degrading enzymes to carbohydrate polymers remaining in pretreated solids. Additionally, pretreatments should not degrade monosaccharides or produce compounds that inhibit downstream processes (e.g. fermentation). They should also require minimal size reduction, consume small amounts of energy, and have low associated costs (Johnson and Elander 2008). Pretreatment technologies can be divided into three main groups— physical, thermochemical, and biological.

2.2.1.1 Physical Pretreatments

Some physical pretreatment is needed for material handling purposes. For agricultural residues and herbaceous energy crops, sufficient size reduction can be achieved with hammer

mills that produce particles able to pass through a ¼-inch screen (Brown 2003). For woody biomass, sufficient size reduction can be achieved with chippers used in the pulp and paper industry (Brown 2003).

Extensive physical pretreatments have also been used, on their own, prior to enzymatic hydrolysis. These include wet, dry, and vibratory ball-milling, as well as compression milling (Alvo and Belkacemi 1997; Millett et al. 1975; Tassinari et al. 1980). Physical pretreatments reduce the size of lignocelluosic particles, increase surface area, and, in some cases, reduce cellulose crystallinity (Fan et al. 1981). However, Mosier et al. (2005) reported that chemical changes that occur in pretreatment reduce cell-wall recalcitrance more effectively than do physical changes. Moreover, it has been shown that mechanical pretreatments, on their own, result in lower yields in enzymatic hydrolysis, use more energy, and have higher costs compared to a number of thermochemical pretreatments (Johnson and Elander 2008).

2.2.1.2 Thermochemical Pretreatments

Thermochemical pretreatment is currently the most promising pretreatment type, both in terms of improving sugar yields through subsequent enzymatic hydrolysis and in terms of cost and energy use (Johnson and Elander 2008; Mosier et al. 2005). Multiple pretreatment processes have been tested on a wide array of lignocellulosic materials, but to date, no one treatment process has been designated as the ideal pretreatment for all biomass types (Chandra et al. 2007; Mosier et al. 2005; Wyman et al. 2009; Wyman et al. 2005b).

Thermochemical pretreatments can be divided into two major classes — acidic and alkaline. Both pretreatment types modify the structure and composition of lignocelluosic

biomass and increase accessibility to cellulase enzymes (Hsu 1996; Mosier et al. 2005; Wyman et al. 2005a; Wyman et al. 2009; Wyman et al. 2005b).

Dilute-acid pretreatments are operated at high temperatures and pressures, and they hydrolyze a significant portion of hemicelluose, chemically alter and redistribute lignin, and leave mostly intact cellulose that is more amenable to enzymatic attack (Hsu 1996). These pretreatments involve addition of acid — usually either H_2SO_4 or SO_2 — or are operated with pure water (e.g. hot water or steam explosion pretreatments). In water-only pretreatments, the acidic catalyst is formed through hydrolysis of hemicellulose acetyl linkages, forming acetic acid. Configurations of dilute-acid treatments include batch (Schell et al. 2003), flow-through (Yang and Wyman 2004), and reactions that involve rapid and explosive decompression [e.g. steam explosion] (Bura et al. 2009; Eklund et al. 1995). The main advantage of acidic pretreatments is their ability to perform well across plant-biomass types (Chandra et al. 2007; McMillan 1994; Nguyen et al. 2000). Disadvantages of acidic pretreatments include the formation of sugar degradation products (e.g. furfural and hydroxymethyl furfural); the corrosive nature of acidic catalysts, which result in high reactor and maintenance costs; and the need to neutralize pretreated substrates prior to enzymatic hydrolysis (Grohmann et al. 1986; Nguyen et al. 2000).

Alkaline pretreatments include the use of sodium hydroxide, calcium hydroxide, ammonia, and ammonium hydroxide, and in contrast to acidic treatments, they can be operated at ambient conditions or at high temperatures and pressures. Alkaline catalysts cleave ester bonds in hemicellulose and lignin-carbohydrate linkages (Chang and Holtzapple 2000; Laureano-Perez et al. 2005; Pavlostathis and Gossett 1985), which, as in the case of acidic treatment, results in a substrate that is more accessible to cell-wall degrading enzymes. Alkaline pretreatments

solubilize a significant fraction of hemicellulose; however, in contrast to acidic treatments, the hemicellulose that remains in the solids is mostly intact, and hence, requires hemicellulases in addition to cellulases for subsequent enzymatic hydrolysis. Chemical and ultrastructural modifications that result from alkaline pretreatments are poorly understood relative to acidic pretreatments (Chundawat 2009). Following are brief descriptions of a number of alkaline pretreatments.

The Ammonia Fiber Explosion (AFEX) process is operated with high concentrations of liquid anhydrous ammonia (1–2 kg ammonia/kg dry biomass) and relatively high solids concentrations of 40% total solids [TS] (Alizadeh et al. 2005; Teymouri et al. 2005). Temperatures range from 60 to 100 ºC, and pressures range from 250 to 300 psig for residence times that range from 5 to 30 minutes, after which the pressure is released rapidly (Alizadeh et al. 2005; Sun and Cheng 2002; Teymouri et al. 2005). AFEX disrupts and rearranges the lignocellulosic matrix, dissolves some lignin and decrystallizes cellulose (Sun and Cheng 2002). This pretreatment has been shown to be effective for agricultural residues and perennial grasses (Alizadeh et al. 2005; Teymouri et al. 2005). However, AFEX does not work well for woody biomasses (Johnson and Elander 2008; McMillan 1994).

Liquid ammonium hydroxide treatments are run at temperatures that range from ambient up to 90 ºC, with chemical loadings of approximately 30 g NH_4OH per 100 g TS, with solids loadings of 10 to 50 % TS, and residence times that range from hours to days (Johnson and Elander 2008). These treatments remove significant amounts of lignin, and, as in the case of AFEX, have been shown to be effective for agricultural residues and perennial grasses (Kim and Lee 2005a; Kim and Lee 2005b). Pretreatments with liquid ammonium hydroxide are operated

in batch (Kim and Lee 2005b; Wyman et al. 2005a) or flow-through mode [e.g. ammonia recycle percolation (ARP)] (Kim and Lee 2005a).

Lime pretreatments have been successfully applied to a variety of feedstocks including corn stover (Karr and Holtzapple 2000), switchgrass (Chang et al. 1997; Karr and Holtzapple 2000), and poplar (Chang et al. 2001; Sierra et al. 2009). Studies on lime pretreatment have been conducted with a number of operating conditions. Temperatures range from ambient to 160 ºC; chemical loadings range from 10 to 40 g $CaOH_2$ per 100 g TS (Chang et al. 1997; Karr and Holtzapple 2000; Sierra et al. 2009). Lime pretreatments are operated with solids loadings of 20% TS or less (Johnson and Elander 2008). Relatively short residence times are employed at higher temperatures, while residence times can extend up to weeks at lower temperatures (Karr and Holtzapple 2000; Sierra et al. 2009). Oxidative conditions are required for effective lime pretreatment of high-lignin, woody biomasses. Blowing air or pure oxygen through pretreatment reactors results in removal of up to 80% of the lignin along with high sugar conversions in subsequent enzymatic hydrolysis of poplar (Chang et al. 2001). One major advantage of using lime is its lower associated cost compared to other alkali chemicals (Johnson and Elander 2008).

Strong sodium hydroxide, high-temperature treatment is widely used in the soda pulping and Kraft pulping processes for delignification of wood. Additionally, sodium hydroxide pretreatments operated at low temperatures have been successfully applied to a number of biomass substrates, including wheat straw (Pavlostathis and Gossett 1985), switchgrass (Xu et al. 2010), and spruce (Zhao et al. 2008). During NaOH pretreatment, alkali is consumed due to saponification of ester bonds that link carbohydrate polymers to each other and to lignin (Pavlostathis and Gossett, 1985). Ester-bond breakage results in solubilization of significant amounts of lignin and hemicellulose, and relatively small amounts of cellulose, resulting in a

substrate enriched in cellulose (Pavlostathis and Gossett 1985; Zhao et al. 2008). Furthermore, NaOH causes substantial swelling of cellulose fibers making them more amenable to enzymatic hydrolysis (Fan et al. 1982; Pavlostathis and Gossett 1985).

Pavlostathis and Gossett (1985) reported that chemical use could be significantly reduced by recovering and reusing unconsumed alkali, without reducing the effectiveness of pretreatment. Through careful water management and use of recycle, NaOH requirements approach alkali consumptive demand. One major advantage of NaOH pretreatment is that it can be operated effectively at ambient temperatures (Pavlostathis and Gossett 1985) Low-temperature treatments minimize cellulose dehydration, which can adversely impact saccharification (Atalla et al. 2008). Moreover, operating at ambient temperatures consumes relatively small amounts of energy, which may contribute to overall lower greenhouse gas emissions from saccharification processes.

2.2.1.3 Biological Pretreatments

Biological pretreatments primarily employ lignin-degrading organisms, such as brown- and white-rot fungi (Keller et al. 2003; Taniguchi et al. 2005). These organisms remove lignin by secreting lignin-degrading enzymes (e.g. lignases, lignin peroxidases, and laccases). With the exception of esterase mixtures, there are no commercially available lignin-active enzyme mixtures. Hence, one drawback of biological pretreatment is loss of carbohydrate, which is degraded by organisms during pretreatment. A second disadvantage of biological delignification is the extremely long retention times — that range from several hours to months — required for satisfactory pretreatment (Johnson and Elander 2008).

Akin et al. (2007) demonstrated that incubating corn stover leaves with a commercial esterase (Depol 740 at loadings of 1 g enzyme per 1 g TS) for 24 h prior to cellulase-mediated hydrolysis increased glucose yields by 1.44 to 2.53 times, compared to cellulase hydrolysis of untreated corn stover leaves. However, treatment with Depol 740 did not increase sugar yields in enzymatic hydrolysis of corn-stover stems (Akin 2007). Since corn stover is a lignocellulosic material with a relatively low level of recalcitrance, and esterase treatment did not improve hydrolysis yields for corn-stover stems, treatment with esterase is not likely to replace thermochemical pretreatment. Furthermore, the enzyme loading of 1 g enzyme per 1 g biomass used in that study is fairly high.

2.2.1.4 Comparison of Leading Thermochemical Pretreatment Technologies

Multiple pretreatment technologies have been characterized and deemed effective at producing substrates with high fermentable sugar yields through subsequent enzymatic hydrolysis. However, the various technologies result in different fractionations of cellulose, hemicellulose, and lignin at different stages of the overall ethanol production process (Chandra et al. 2007; Mosier et al. 2005; Wyman et al. 2005b; Wyman et al. 2009). Additionally, the chemical and physical mechanisms by which thermochemical pretreatments improve enzymatic saccharification vary and are not well understood (Chandra et al.2007; Mosier et al. 2005; Wyman et al. 2005b; Wyman et al. 2009). Examples of changes in physical and chemical characteristics that have been cited as contributing to pretreatment effectiveness include increased porosity (Grethlein 1985); reduced crystallinity and degree of polymerization of cellulose (Sun and Cheng 2002; Fan et al. 1991); and removal of hemicellulose (Ohgren 2007)

and lignin (Grabber 2005). The observed effects of selected, leading pretreatment technologies on cellulose, hemicellulose, and lignin are summarized in Table 2.1.

Table 2.1 Changes is biomass compositional features for leading pretreatment technologies. Adapted from Zhu 2004 and Mosier et al. 2005.

	Compositional features			
Pretreatment	**Cellulose**	**Hemicellulose**	**Lignin**	**References**
Dilute-acid	Some depolymerization	80-100% solubilization	Little solubilization; redistribution	Grethlein and Converse 1991
Liquid Hot Water	ND	~ 80% solubilization	Little solubilization; redistribution	Mosier et al. 2005
Steam Explosion	Some depolymerization	80-100% solubilization	Little solubilization; redistribution	Grethlein and Converse 1991
AFEX	Decrystalization	Up to 60% solubilization	10-20% solubilization	Dale and Moreira 1982
ARP	Less than 5% depolymerization	~ 50% solubilzation	~ 70% solubilization	Kim and Lee 2005a; Kim and Lee 2005b
Lime	Little depolymerization	~ 30% solubilization	~ 40% solubilization	Chang et al. 1997
Sodium hydroxide	Substantial swelling; Type I → Type II	Up to 80% solubilization	~ 50% solubilization	Millet et al. 1975; Pavlostathis and Gossett 1985

2.2.1.5 Future Prospects

Pretreatment is associated with significant costs (Johnson and Elander 2008), and is hence worthy of research and development. Moreover, integrated studies that measure pretreatment effects while considering both upstream processes (e.g. harvest and storage), as well as downstream processes (e.g. fermentation) are needed. Researchers have proposed that due to the wide heterogeneity of lignocellulosic biomass, pretreatment processes will need to be tailored specifically to different biomasses (Chandra et al. 2007; Merino and Cherry 2007; Mosier et al. 2005). This has the potential to create cumbersome logistics for a prospective technology that is expected to involve the application of varying, seasonal energy crops to a single, lignocellulose-to-ethanol facility. Thus, more studies should be conducted to measure affects of biomass variation on the design of pretreatment processes.

2.2.2 Enzymatic Hydrolysis

Microbial cell-wall-degrading enzymes involved in the deconstruction of pretreated lignocellulosic substrates include cellulases, hemicellulases, and esterases. In the last decade, the cost of industrial cellulase cocktails was reduced by about a factor of ten from a starting cost of $5 per gallon of ethanol produced (Himmel and Picataggio 2008). However, this cost reduction is not sufficient to make cellulosic biorefinery products, such as ethanol and butanol, cost-competitive with petroleum-based fuels or starch-based biofuels (Himmel and Picataggio 2008). High concentrations of enzymes (typically 25-40 mg per g TS) and long retention times (typically 72 h) are still required for effective hydrolysis of pretreated lignocellulosic substrates.

Cellulases are enzymes that catalyze hydrolysis of β-1,4 glycosidic bonds that link glucose residues in cellulose. There are three major types of cellulases: endocellulases that

cleave internal glycosidic bonds, mostly in regions of amorphous cellulose; exocellulases that hydrolyze glycosidic bonds at the ends of glucose chains and release cellobiose; and processive endocellulases that act on both amorphous and crystalline cellulose (Wilson 2008). Cellulases consist of at least two domains connected by a flexible glycosylated linker region (Wilson 2008). A catalytic domain contains the enzyme active site, and a cellulase binding module (CBM) plays a major role in enzyme binding, a prerequisite for hydrolysis of crystalline cellulose (Wilson 2008). An additional enzyme β-glucosidase is required to hydrolyze cellobiose into glucose.

Cellulolytic organisms are diverse and they include aerobic and anaerobic microorganisms. Industrial enzymes expected to be used in biorefineries, however, come from aerobic fungi. *Trichoderma reesei*, also referred to as *Hypocrea jecorina*, has been engineered to secrete high concentrations of protein with high specific activity on cellulose (Ghosh et al. 1984). The cellulolytic enzyme system of *T .reesei* is composed of two exocellulases, seven endocellulases, and two β-glucosidases (Markov et al. 2005) — although *T. reesei* cocktails are typically deficient in ß-glucosidase activity. Cel7A, an exocellulase that cleaves cellobiose from the reducing end of cellulose chains, and Cel6A, an exocellulase that cleaves cellobiose from the non-reducing end of cellulose chains, make up 70% and 10 % of *T. reesei's* cellulase protein, respectively (Markov et al. 2005). In addition to being the most abundant cellulase secreted by *T. reesei*, Cel7A also has the highest specific activity and plays a major role in crystalline cellulose hydrolysis. One might assume that *T. reesei* secretes nine cellulases for reasons that have to do with synergism. However, cellulase synergism is only observed when enzymes with different activities (e.g. endocellulase and exocellulase) degrade crystalline cellulose (Wilson 2008). Thus, it is surprising that *T. reesei* secretes seven different endocellulases, since the endocellulases do not show synergism with one another (Wilson 2008).

Hemicellulases, in contrast to cellulases, vary in the types of bonds they act on. Hydrolysis of hemicellulose requires multiple enzyme activities (Saha 2003), including glycosyl hydrolases that cleave various glycosidic bonds within carbohydrate polymers and esterases that cleave ester bonds that link carbohydrate polymers to each other and to lignin (Decker et al. 2008). For example, enzymatic saccharification of arabinoxylan (Fig. 2.2) requires activities that depolymerize the xylan backbone, as well as multiple side-group cleaving activities. Xylan-backbone-hydrolyzing enzymes include exo- and endo-ß-1,4-xylanases; side-group-cleaving enzymes include α-L-arabinofuranosidases, feruloyl esterases, *p*-coumaryl esterases, and acetyl esterases; and ß-xylosidase is required to hydrolyze xylobiose into xylose (Decker et al. 2008; Gilbert et al. 2008; Saha 2003; Sørensen et al. 2007).

Enzymatic breakdown of hemicellulose has not been studied as extensively as cellulose hydrolysis (Gilbert et al. 2008; Sørensen et al. 2005). The focus of researchers on cellulases is due, in part, to wide use of dilute-acid pretreatments, which remove most of the hemicellulose prior to enzymatic hydrolysis (Gilbert et al. 2008). However, several researchers have reported that hemicellulase supplementation of cellulase-mediated hydrolysis of multiple feedstocks that underwent acidic or alkaline pretreatments resulted in increased glucose production (Berlin et al. 2007; Bura et al. 2009; Kumar and Wyman 2009a; Kumar and Wyman 2009b; Murnen et al. 2007; Ohgren et al. 2007; Zhong et al. 2009). It is surprising that hemicellulase supplementation of acid-pretreated biomasses would result in increased glucose yields, but researchers have demonstrated this for a number of substrates (Berlin et al., 2007; Bura et al., 2009; Kumar and Wyman 2009a, 2009b; Ohgren et al., 2007). Alkaline pretreatments produce solids that retain the majority of hemicellulose polymers intact, and as expected, hemicellulase supplementation increased glucose and xylose yields for numerous substrates (Bura et al. 2009; Kumar and

Wyman 2009a; Kumar and Wyman 2009b; Murnen et al. 2007; Prior and Day 2007; Zhong et al. 2009).

2.3 Fourier Transform Infrared Spectroscopy (FTIR)

Standard wet-chemistry analyses of lignocellulosic biomass [e.g. NREL two-step, acid-hydrolysis protocol (Sluiter et al. 2008), and the acid-detergent method (Van Soest et al. 1991)] quantify glucan (or cellulose), xylan (or hemicellulose), and lignin. However, they are labor-intensive and do not differentiate among types of cellulose (e.g. I or II), hemicellulose (e.g. xyloglucan or arabinoxylan) or lignin (e.g. syringyl or guaicyl), and do not provide insight into chemical bonds that link these three components to one another. Researchers have demonstrated that the forms of cellulose (Atalla et al. 2008), hemicellulose (Gilbert et al. 2008; Saha 2003), lignin (Grabber 2005), and carbohydrate-lignin bonds (Casler and Jung 2006) affect recalcitrance and, in turn, sugar yields from pretreatment enzymatic hydrolysis. Therefore, there is need for a method that characterizes lignocellulose based upon chemical attributes associated with recalcitrance, allowing prediction of sugar production attainable from pretreated biomass without having to conduct trials of enzymatic hydrolysis — or perhaps without having to conduct wet-chemistry analyses at all.

One of the objectives of this dissertation is to ascertain the potential for using FTIR, combined with partial least squares (PLS) regression modeling, to replace wet-chemistry analyses and enzymatic assays in estimating saccharification from various biomasses.

2.3.1 Principles of Operation

An infrared (IR) spectrum is obtained when a sample absorbs radiation in the IR region of the electromagnetic spectrum. The mid-IR region includes wavelengths of 2.5 to 25 µm, which correspond to wavenumbers of 4000 to 400 cm^{-1} (Colthup et al. 1964). For organic molecules, vibrational modes that correspond to, and, thus, absorb IR radiation include stretching (symmetric and asymmetric) and bending (rocking, scissoring, and wagging) vibrations that result from dipole moments of asymmetric covalent bonds (Colthup et al. 1964; Williams and Fleming 1980). Particular chemical bonds in organic samples absorb radiation of specific frequencies, allowing "fingerprinting" of organic samples, such as raw and pretreated plant biomasses (McCann et al. 1997). The more asymmetric covalent bonds present in a sample, the more complex the resulting IR spectrum.

The following description of the operating principles of an FTIR instrument was adapted from Perkins (1987) and McCann (1997). FTIR spectroscopy consists of two steps. In the first step, an interferogram is collected with a Michelsen interferometer, and in the second step the interferogram is mathematically transformed into a spectrum using a Fourier transform algorithm (Cooley and Tukey 1965). A simple Michelsen interferometer consists of an IR source, a beam splitter, two mirrors located at equal distances from the sample port, a reference beam, and a detector. The IR source is split in two with a beam splitter. One beam is directed towards a stationary mirror, while the second beam is directed towards a moving mirror. The mirrors reflect the two IR beams to meet back at the beam splitter, which directs the combined beams towards the sample. The radiation not absorbed by the sample is then transmitted to and collected by a detector. The moving mirror produces varying optical path differences between the two beams, resulting in constructive and destructive interference that varies with the

translation of the moving mirror. A reference beam with a known wavenumber is used to precisely measure the displacement of the moving mirror. The result is an interferogram, which is absorbance data as a function of distance. The interferogram is then mathematically transformed with a Fourier transform algorithm (Cooley and Tukey 1965), resulting in a spectrum, which is absorbance data as a function of wavenumber in units of cm^{-1}.

FTIR spectrometers can be operated in three major modes: transmittance, diffuse reflectance, and attenuated total reflectance [ATR] (PerkinElmer Life and Analytical Sciences 2005). FTIR-ATR analysis can accommodate relatively thick solid samples, requires minimal sample preparation, and exhibits superior reproducibility compared to transmittance and diffuse reflectance modes when applied to lignocellulosic samples (Zhou et al. 2011).

In FTIR-ATR spectroscopy, infrared radiation is passed through a diamond or germanium crystal and is reflected off the area of the crystal that is in contact with a solid sample, forming an evanescent wave (PerkinElmer Life and Analytical Sciences 2005). Solid samples must be in contact with the ATR crystal for the evanescent wave to form. The evanescent wave then penetrates into the sample and non-absorbed radiation is passed onto a detector (PerkinElmer Life and Analytical Sciences 2005). The depth of penetration of the evanescent wave varies from 0.5 to 2 μm, depending on the angle of incidence as well as the refractive indices of the ATR crystal and the sample (PerkinElmer Life and Analytical Sciences 2005).

2.3.2 Spectral Preprocessing

Mathematical preprocessing of FTIR spectra prior to data analysis is recommended (Wehrens 2011). Spectral treatments, including normalization, baseline correction,

multiplicative scatter correction, smoothing, and differentiation, have been shown to improve subsequent analyses using FTIR spectra (Moros et al. 2010). Within-spectra normalization can account for differences in sample thickness, refractive indices, orientation, and effects of light scattering (Nicolai et al. 2007; Wehrens 2011). Additionally, FTIR-ATR spectroscopy requires a correction to account for varying depth of penetration of the IR radiation into the sample across wavenumbers (PerkinElmer Life and Analytical Sciences 2005).

2.3.3 FTIR Spectroscopy Applied to Lignocellulose

The complexity of lignocellulosic substrates results in complicated FTIR spectra with overlapping bands. Given these limitations, it is surprising that there are a number of researchers who have used FTIR to characterize isolated cellulose, hemicellulose, and lignin. Nelson and O'Connor (1964) measured crystallinity of pure cellulose, by calculating the ratio of absorbance intensities at 1430 and 890 cm^{-1}. Since then, this ratio, referred to as the Total Crystallinity Index, has been correlated to other, accepted measurement techniques that quantify cellulose crystallinity, such as Nuclear Magnetic Resonance (NMR) and X-ray diffraction (Klemm et al. 2005; Sugiyama et al. 1991). Robert et al. (2005) demonstrated that FTIR could elucidate structural details of previously characterized arabinoxylan polymers with different side chains and degrees of substitution. Collier et al. (1992) used isotope labeling to demonstrate it is possible to assign IR absorbance bands to aryl-alkyl C-O bonds in model lignin compounds.

Recently, Corgié et al. (2011) used FTIR to measure extents of enzymatic hydrolysis of pure cellulose, correlating sugar yields directly to previously assigned cellulose signals. Details on changes in the structure of crystalline cellulose resulting from separate and synergistic activities of three pure cellulases from the bacterium *Thermobifida fusca* were revealed (Corgié

et al. 2011). The enzymes used in that study, a classical endocellulase (Cel5A), a classical exocellulase (Cel6B), and a processive endocellulase (Cel9A), impacted the cellulose structure differently. While Cel9A preferentially degraded the I_{α} allomorph of cellulose, Cel5A and Cel6B, acting together, preferentially degraded the $I_{ß}$ allomorph of cellulose, which has been shown to be the most difficult cellulose allomorph to degrade (Qiang et al. 2005). Surprisingly, changes in hydrogen bonding patterns did not correlate with sugar production, suggesting that FITR may help elucidate mechanisms by which cellulases or CBMs alter H-bonding patterns in cellulose without cleaving ß-glycosidic bonds (Corgié et al. 2011).

Unfortunately, the complexity of intact lignocellulose and the similarity of bonds and functional groups in cellulose, hemicellulose, and lignin result in IR spectra that cannot be interpreted straightforwardly. Qualitative studies have been conducted where relative changes in absorbance intensities of different FTIR bands were monitored during chemical and biological processing of lignocellulosic substrates. Pandey and Pitman (2003) monitored biodegradation of wood by white- and brown-rot fungi by measuring relative changes in absorbance intensities associated with carbohydrate and lignin. Additionally, a number of studies provided qualitative descriptions of relative increases and decreases in FTIR absorbances to help explain results of studies on thermochemical pretreatment of lignocellulosic substrates (Chundawat et al. 2007; Guo et al. 2009; Kristensen et al. 2008).

Multivariate statistical models, such as principal components (PC) and partial least squares (PLS) regression, have been applied to FTIR spectra of lignocellulosic substrates to conduct quantitative compositional analyses. PC regression models combined with FTIR have been used to accurately quantify lignin in wood (Raiskila et al. 2007). PLS regression models combined with FTIR have been use to accurately predict lignin, hydroxycinnamic acids, and

protein in two perennial grasses (Allison et al. 2009a; Allison et al. 2009b); cellulose, hemicellulose, and lignin in raw (Chen et al. 2010) and pretreated (Tucker et al. 2001) wood; carbohydrate and lignin in wood biodegraded by white- and brown-rot fungi (Ferraz et al. 2000); carbohydrate, ash, extractives, and lignin in straw (Tamaki and Mazza 2011a; Tamaki and Mazza 2011b); and lignin and energy content in poplar (Zhou et al. 2011). Such models were constructed by correlating FTIR spectra to the chemical constituent of interest as measured by wet-chemistry techniques.

To the author's knowledge, only one published study exists where multivariate regression was applied to FTIR spectra of pretreated biomass in an attempt to estimate saccharification from enzymatic hydrolysis. (Gollapalli et al. 2002) found that FTIR combined with PC regression did not adequately predict glucose and xylose yields from hydrolysis of rice straw that had been pretreated by AFEX. However, it is not clear to what extent they explored alternative techniques for normalization of FTIR spectra and for removing unimportant latent variables, techniques that can improve PC regression models (Raiskila et al. 2007).

2.4 Chemometrics — Multivariate Data Analysis

Vibrational spectroscopy techniques such as FTIR, Fourier transform near infrared (FTNIR), and NMR present researchers with large and complex data sets, which are difficult to interpret on their own. Furthermore, spectroscopic analyses result in high dimensional data with large numbers of variables and relatively limited sample sizes. For example, one FTIR spectrum may consist of thousands of absorbance intensities. Additionally, multicollinearity often exists in such data sets, causing numerical and statistical problems with traditional statistical methods

such as Ordinary Least Squares (OLS) regression. Chemometrics emerged as a way to model data for which OLS regression is not appropriate (Eriksson 2001; Wehrens 2011).

To illustrate this problem, consider a matrix of predictor variables, $\boldsymbol{X}[n \times k]$, with dimensions of n observations by k variables, and a dependent variable, $\boldsymbol{Y}[n \times 1]$, with dimensions of n observations by 1 dependent variable. If the goal is to predict $\boldsymbol{Y}$ from $\boldsymbol{X}$, the simplest method to use is ordinary least squares regression. The model is given by

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{e} \qquad (2.1)$$

where $\boldsymbol{B}[k \times 1]$ is a regression coefficient matrix with dimensions of k predictor variables by 1 dependent variable, and $\boldsymbol{e}[n \times 1]$ are residuals. If the $\boldsymbol{X}$ matrix is of full rank, the regression coefficient matrix, $\boldsymbol{B}$, can be estimated with the least squares estimator

$$\hat{\boldsymbol{B}} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y} \qquad (2.2)$$

where $\hat{\boldsymbol{B}}$ is the estimate of $\boldsymbol{B}$, and the superscript T represents transpose. However, when $k >> n$, the $\boldsymbol{X}$ matrix may be singular, which will result in parameter estimates, $\hat{\boldsymbol{B}}$, that are not unique. Hence, Eq. (2.2) cannot be solved. One solution to this problem is to reduce the number of variables through latent variable projection. The most popular technique for variable reduction, in this situation, is principal components analysis (Jolliffe 2002).

2.4.1 Principal Components Analysis (PCA)

The goal of PCA is to compress the data matrix, $\boldsymbol{X}[n \times k]$, into fewer independent (orthogonal) latent variables, or principal components, that capture the variation in the original data (Jolliffe 2002). These new latent variables are eigenvectors extracted from the $\boldsymbol{X}$ matrix,

which may be calculated using either the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm (Höskuldsson 1995; Wold 1966) or singular value decomposition [SVD] (Jolliffe 2002), to obtain *z* latent variables or principal components.

Assuming the number of variables is reduced from *n* to *z*, the new latent variables, or principal components, are comprised of scores, $\boldsymbol{T}[n \times z]$, and loadings, $\boldsymbol{P}[k \times z]$, that capture most of the variation contained in $\boldsymbol{X}$. The resulting PCA model describes the original data matrix $\boldsymbol{X}[n \times k]$ as the sum of extracted components, namely

$$\hat{\boldsymbol{X}} = \boldsymbol{T}\boldsymbol{P}^T = \boldsymbol{t}_1\boldsymbol{p}_1^T + \boldsymbol{t}_2\boldsymbol{p}_2^T + \ldots + \boldsymbol{t}_z\boldsymbol{p}_z^T \qquad (2.3)$$

where $\boldsymbol{t}_z$ is the z^{th} column vector of $\boldsymbol{T}[n \times z]$ and $\boldsymbol{p}_z^T$ is the transpose of the z^{th} column vector of $\boldsymbol{P}[k \times z]$. PCA scores and loadings can be interpreted to see if there are groupings among observations and/or among samples, and PCA can be used to identify outliers in large sample matrices (Joliffe 2002). Additionally, PC regression models can be built by correlating the reconstructed $\boldsymbol{X}$-matrix, $\hat{\boldsymbol{X}}$ $[n \times z]$ in Eq. (2.3) to the dependent variable $\boldsymbol{Y}[n \times 1]$ using ordinary least squares regression. However, $\hat{\boldsymbol{X}}$ $[n \times z]$ is extracted from the original data by calculating the first few eigenvectors that capture the maximum variation in $\boldsymbol{X}$, without considering the dependent variable, $\boldsymbol{Y}$. It would be advantageous to have a data compression method and construct a regression model that correlates $\boldsymbol{X}$ and $\boldsymbol{Y}$, using latent variables extracted from $\boldsymbol{X}$ that capture the maximum covariance of $\boldsymbol{X}$ and $\boldsymbol{Y}$. PLS regression does just that (Eriksson et al. 2001; Geladi and Kowalski 1986; Höskuldsson 1988; Wold et al. 2001.)

2.4.2 Partial Least Squares (PLS) Regression

PLS stands for "Projections to Latent Structures" by means of partial least squares (Wold et al. 2001), and its usefulness lies in its ability to analyze large collinear data sets. PLS regression can relate two data matrices, $\boldsymbol{X}[n \times k]$ and $\boldsymbol{Y}[n \times 1]$ using Eq. (2.1), when $n >> k$. Like PCA, PLS regression compresses $\boldsymbol{X}[n \times k]$ into a smaller number of latent variables that capture most of the data variation. However, in contrast to PCA, the latent variables in PLS regression are chosen to maximize the covariance of $\boldsymbol{X}[n \times k]$ and $\boldsymbol{Y}[n \times 1]$, where $\text{Cov}(\boldsymbol{XY}) = \boldsymbol{XY}^T$.

Following is a brief overview of the linear PLS model, as described by Wold et al. (2001). Detailed descriptions of the PLS model are presented in Eriksson et al. (2001), Geladi and Kowalski (1986), Höskuldsson (1988), and Wold et al. (2001).

A linear PLS regression model with z latent variables is given by

$$\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{F} \qquad (2.4)$$

where $\boldsymbol{B}[k \times 1]$ is the regression coefficient matrix and $\boldsymbol{F}[n \times 1]$ are residuals. The linear PLS regression model finds z new latent variables, which are called $\boldsymbol{X}$-scores, and are denoted by $\boldsymbol{T}[n \times z]$. The $\boldsymbol{X}$-scores, $\boldsymbol{T}[n \times z]$, model $\boldsymbol{X}[n \times k]$ and predict $\boldsymbol{Y}[n \times 1]$ — i.e. $\boldsymbol{Y}$ and $\boldsymbol{X}$ are assumed to be modeled by the same latent variables. The $\boldsymbol{X}$-scores, $\boldsymbol{T}[n \times z]$, are estimated as linear combinations of the original variables, $\boldsymbol{X}$, with $\boldsymbol{X}$-weights, $\boldsymbol{W}[k \times z]$

$$\boldsymbol{T} = \boldsymbol{XW} \qquad (2.5)$$

Furthermore, the $\boldsymbol{X}$-scores, $\boldsymbol{T}[n \times z]$, are multiplied by the $\boldsymbol{X}$-loadings, $\boldsymbol{P}[k \times z]$, such that the residuals, $\boldsymbol{E}[n \times k]$, are "small" (Wold et al. 2001), so that

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^T + \boldsymbol{E} \qquad (2.6)$$

The dependent variable, $\boldsymbol{Y}$, is decomposed into $\boldsymbol{Y}$-scores, $\boldsymbol{U}[n \times z]$, and $\boldsymbol{Y}$-weights, $\boldsymbol{C}[1 \times z]$ such that the residuals, $\boldsymbol{G}[n \times 1]$, are "small" (Wold et al. 2001), namely

$$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{C}^T + \boldsymbol{G} \qquad (2.7)$$

Additionally, the $\boldsymbol{X}$-scores, $\boldsymbol{T}[n \times z]$, are chosen to be good predictors of $\boldsymbol{Y}$, such that residuals, $\boldsymbol{F}[n \times 1]$, are "small" (Wold et al. 2001):

$$\boldsymbol{Y} = \boldsymbol{T}\boldsymbol{C}^T + \boldsymbol{F} \qquad (2.8)$$

A number of algorithms may be used to calculate $\boldsymbol{W}[k \times z]$ and $\boldsymbol{C}[1 \times z]$, such that the covariance of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is maximized (Geladi and Kowalski 1986; Höskuldsson 1988, and Wold et al. 2001).

Substituting Eq. (2.5) into Eq. (2.8) yields

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{W}\boldsymbol{C}^T + \boldsymbol{F} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{F} \qquad (2.9)$$

where

$$\boldsymbol{B} = \boldsymbol{W}\boldsymbol{C}^T \qquad (2.10)$$

2.4.2.1 Model Construction

Prior to building a PLS regression model, $\boldsymbol{X}$ and $\boldsymbol{Y}$ variables must be mean-centered (for $\boldsymbol{X}$, this means that each column is mean-centered), and in some cases scaled. Often, when predictor variables (columns of $\boldsymbol{X}$) are of greatly different magnitudes, variables are auto-scaled,

making all variables have a mean of zero and a variance of one (Eriksson et al. 2001; Geladi and Kowalski 1986; Wold et al. 2001). However, auto-scaling is not recommended for spectral data because it causes spectral noise and actual peaks to be of equal magnitudes (Eriksson et al. 2001; Wehrens 2011). Furthermore, if a data set consists of two types of variables, and each have vastly different numbers of variables, "block scaling" should be conducted (i.e. each variable type should be scaled differently), such that the model accurately represents the importance of both variable types (Eriksson et al. 2001). Ultimately, when predictor variables are of different magnitudes, it is up to the modeler to determine appropriate scaling, and more often than not, proper scaling is determined via trial and error (Eriksson et al. 2001; Wold et al. 2001).

In addition to scaling, transformation of variables is sometimes necessary to accommodate nonlinear relationships between predictor and response variables (Eriksson et al. 2001; Wold et al. 2001). Transformations used in linear PLS models include log, power, and reciprocal transformations, and, as in the case of variable scaling, an appropriate variable transformation is often determined via trial and error (Eriksson et al. 2001; Wold et al. 2001).

The SIMPLS (de Jong 1993) and orthogonal scores [OSCORES] (Martens and Naes 1989) or NIPALS (Wold 1966) algorithms were used to fit PLS models in Chapters 4 and 5.

2.4.2.2 Model Selection and Assessment

Constructing PLS regression models requires the practitioner to choose the number of latent variables or components retained in the model. Retaining more components results in a better fit of the calibration data but raises the risk of over-fitting, which results in models with poor predictive power. Model validation is used to choose "the best" number of components that

have good balance between model fit and ability to predict new samples (Eriksson, 2001; Geladi and Kowalski 1986; Wold et al. 2001).

When a data set is large enough, it is best to divide the data into two groups and build a model with one data set, called a "training set," and test the predictive ability of the model with a second data set, called a "test set" (Eriksson 2001). One needs approximately 20 samples for each data set, and more samples are recommended (Wold et al. 2001). However, when the data set is not sufficiently large to divide in two, cross-validation is a reasonable alternative (Eriksson 2001; Höskuldsson 1988; Wold 2001).

Leave-one-out (LOO) cross-validation is a technique where each sample is removed in turn from the data set, and then predicted by inputting its corresponding row from $\boldsymbol{X}[n \text{ x } k]$, $\boldsymbol{x_i}[1 \text{ x } k]$, into the PLS model built without $\boldsymbol{x_i}$ and y_i (Stone 1974; Wehrens 2011). For each sample, an error is calculated as the difference between $\hat{y}_i$ and y_i where $\hat{y}_i$ is the estimate of y_i predicted from the PLS model constructed without $\boldsymbol{x_i}$ and y_i. This error is calculated for each of the n observations to obtain a root mean square error of cross validation (RMSECV), which is defined as

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2}{n}} \qquad (2.11)$$

where y_i is an element of the vector $\boldsymbol{Y}[n \text{ x } 1]$ and n represents the number of samples (or observations). Relatively small values of RMSECV indicate PLS models with good predictive power.

An additional diagnostic measure for cross-validation is the coefficient of determination for cross-validation, Q^2, which is defined as

$$Q^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \qquad (2.12)$$

where $\bar{y}_i$ is the mean of the response values not including y_i. Note that the maximum value of Q^2 is unity. High values of Q^2 indicate PLS models with good predictive power. Typically the magnitude of RMSECV initially decreases with increasing numbers of latent variables, and as over-fitting becomes a problem, values of RMSECV increase. Similarly, values of Q^2 initially increase as latent variables are added to the model until a maximum is reached, after which values of Q^2 decrease, again indicating the presence of over-fitting. The number of latent variables that minimizes RMSECV and maximizes Q^2 is chosen for model construction.

Once the number of latent variables included in the PLS regression model is chosen, the root mean square error of calibration (RMSEC) and coefficient of determination for calibration, R^2, may be calculated as

$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \qquad (2.13)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad (2.14)$$

where $\hat{y}_i$ is the model fit for y_i with a model constructed from all *n* samples, and $\bar{y}$ is the mean of all *n* response values. Note that the maximum value of R^2 is unity.

Reports in the literature claim that models with Q^2 values between 0.80 and 0.90 have good predictive power and values higher than 0.90 indicate excellent predictive power (Tamaki and Mazza 2011b). The quantity R^2 is a measure of how well the model fits the present data, and it is not recommended as a measure of how well the model might predict with new data. However, large differences between Q^2 and R^2 indicate over-fitting and it has been suggested that the difference should be less than 0.2 for models with good predictive ability (Chen et al. 2010). In the end, of course, the usefulness of any model is determined by whether its prediction error is acceptable for its intended use — a determination that is inevitably context-dependent.

2.4.2.3 *Variable Selection*

The aim of variable selection is to remove predictor variables that do not explain the dependent variable. Removing variables has been shown to improve the predictive ability of PLS regression models while making them simpler (i.e. lower number of latent variables) and thus easier to interpret (Andersen and Bro 2010; Chong and Jun 2005; Wold et al. 2001). Two variable selection methods are used in Chapters 4 and 5. In the first method, 90% confidence intervals (CIs) of the regression coefficients [$\boldsymbol{B}$ in Eqs. (2.9) and (2.10)] are calculated; the second method uses variable important for projection (VIP) scores and is referred to in the literature as the VIP method (Chong and Jun 2005).

With the first variable selection method, regression coefficients with 90% CIs that span the value of zero are considered unimportant for predicting $\boldsymbol{Y}$, and their associated variables are

removed from the regression model (Wehrens 2011). Since there are no analytical solutions for standard errors of PLS regression coefficients, 90% (CIs) of the regression coefficients may be calculated using a nonparametric bootstrapping technique with a percentile method (Efron and Tibshirani 1986; Wehrens 2011). Briefly, two thousand bootstrap samples are generated, and each is a subset of the data, where some samples from the data set are present multiple times, while others are absent, and regression coefficients are calculated for each bootstrap model. The 5th and 95th percentiles of each regression coefficient are then used to define 90 % CIs.

The criterion used in the VIP method is that variables with VIP scores less than 1 are considered unimportant and removed from the model (Chong and Jun 2005). A VIP score describes the importance of the k^{th} variable for explaining $\boldsymbol{X}$ and for predicting the dependent variable, $\boldsymbol{Y}$ (Anderson and Bro, 2010; Chong and Jun, 2005). For a model with z latent variables, this score is given by

$$\mathrm{VIP}_k = \sqrt{\frac{k\sum_{i=1}^{z}\left(\mathrm{SS}\left(b_i\boldsymbol{t_i}\right)\left(w_{ki}/\|\boldsymbol{w_i}\|\right)^2\right)}{\sum_{i=1}^{z}\mathrm{SS}\left(b_i\boldsymbol{t_i}\right)}} \qquad (2.15)$$

where $\mathrm{SS}\left(b_i\boldsymbol{t_i}\right) = b_i^2\boldsymbol{t_i}^T\boldsymbol{t_i}$ and $b_i = \boldsymbol{Y_i}^T\boldsymbol{t_i} / \boldsymbol{t_i}^T\boldsymbol{t_i}$. Note that VIP_k is the VIP score for the k^{th} predictor variable; $\boldsymbol{t_i}$ is the i^{th} column vector of the $\boldsymbol{X}$-scores, $\boldsymbol{T}[n \times z]$; w_{ki} is the weight for the k^{th} variable of a model with i latent variables; $\boldsymbol{w_i}$ is the i^{th} column vector of the $\boldsymbol{X}$-weights, $\boldsymbol{W}[k \times z]$; and $\boldsymbol{Y_i}$ is the deflated $\boldsymbol{Y}$ vector after the model fit with $i - 1$ latent variables is subtracted from $\boldsymbol{Y}$. Note that calculation of VIP scores requires fitting the PLS model with the NIPALS algorithm (Wold 1966) or the equivalent OSCORES algorithm (Martens and Naes 1989).

Chong and Jun (2005) compared two variable-selection techniques with a simulated data set. The first variable selection technique described in that study used relative magnitudes of the regression coefficients to identify insignificant variables. The second method defined insignificant variables as those in which $VIP_k < 1$. The authors reported that the two variable selection methods were complementary. The VIP method worked better for autoscaled data with regression coefficients of similar magnitudes, and the method that compared magnitudes of regression coefficients worked better for data sets with variables and regression coefficients of different magnitudes (Chong and Jun 2005). Additionally, Wold et al. (2001) reported that a variable should be removed from PLS regression models, only if the magnitude of its regression coefficient is very small and its VIP score is less than 1. Ultimately, models constructed after variable selection, should be tested using the model assessment described in Section 2.4.2.2.

REFERENCES

Akin DE. 2007. Grass Lignocellulose. Appl Biochem and Biotechnol 136-140:3-15.

Alizadeh H, Teymouri F, Gilbert T, Dale B. 2005. Pretreatment of switchgrass by ammonia fiber explosion (AFEX). Appl Biochem Biotechnol 124:1133-1141.

Allison GG, Thain SC, Morris P, Morris C, Hawkins S, Hauck B, Barraclough T, Yates N, Shield I, Bridgwater A and others. 2009a. Measurement of key compositional parameters in two species of energy grass by Fourier transform infrared spectroscopy. Bioresour Technol 100:6428-6433.

Allison GG, Thain SC, Morris P, Morris C, Hawkins S, Hauck B, Barraclough T, Yates N, Shield I, Bridgwater A and others. 2009b. Quantification of hydroxycinnamic acids and lignin in perennial forage and energy grasses by Fourier-transform infrared spectroscopy and partial least squares regression. Bioresour Technol 100:1252-1261.

Alvo P, Belkacemi K. 1997. Enzymatic saccharification of milled timothy (*Phleum pratense L.*) and alfalfa (*Medicago sativa L.*). Bioresour Technol 61:185-198.

Andersen CM, Bro R. 2010. Variable selection in regression — a tutorial. J Chemometrics 24:728-737.

Atalla RH, Brady JW, Matthews JF, Ding S-Y, Himmel ME. 2008. Structure of plant cell wall celluloses. In: Himmel ME, editor. Biomass recalcitrance: deconstructing the plant cell wall for bioenergy. Oxford, United Kingdom: Blackwell Publishing Ltd. p 188-212.

Atalla RH, Hackney J, Uhlin I, Thompson N. 1993. Hemicelluloses as structural regulators in the aggregation of native cellulose. Intl J Biol Macromol 15:109-112.

Berlin A, Balakshin M, Gilkes N, Kadla J, Maximenko V, Kubo S, Saddler J. 2006. Inhibition of cellulase, xylanase and β-glucosidase activities by softwood lignin preparations. J Biotechnol 125:198-208.

Berlin A, Maximenko V, Gilkes N, Saddler J. 2007. Optimization of enzyme complexes for lignocellulosic hydrolysis. Biotechnol Bioeng 77:287-295.

Besle J, Cornu A, Jouany J. 1994. Roles of structural phenylpropanoids in forage cell wall digestion. J Sci Food Agric 64:171-190.

Brown RC. 2003. Biorenewable resources - engineering new products from agriculture. Ames, Iowa: Wiley-Blackwell.

Brown Jr. RM, Saxena, IM, Kudlicka. 1996. Cellulose biosynthesis in higher plants. Trends Plant Sci 1:149-156.

Bura R, Chandra RP, Saddler J. 2009. Influence of xylan on the enzymatic hydrolysis of steam-pretreated corn stover and hybrid poplar. Biotechnol Prog 25:314-322.

Carpita NC, McCann MC. 2000. The cell wall. In: Buchanan BB, Gruissem W, Jones RL, editors. Biochemistry and Molecular Biology of Plants. Rockville, MD: Am Soc Plant Biol. p 281-309.

Casler MD, Jung H-JG. 2006. Relationships of fibre, lignin, and phenolics to in vitro fibre digestibility in three perennial grasses. Anim Feed Sci Technol 125:151-161.

Chandra RP, Bura R, Mabee WE, Berlin A, Pan X, Saddler JN. 2007. Substrate pretreatment: the key to effective enzymatic hydrolysis of lignocellulosics? Adv Biochem Eng Biotechnol 108:67-93.

Chang VS, Burr B, Holtzapple MT. 1997. Lime pretreatment of switchgrass. Appl Biochem Biotechnol 63-65:3-19.

Chang VS, Holtzapple M. 2000. Fundamental factors affecting biomass enzymatic reactivity. Appl Biochem Biotechnol 84-86:5-37.

Chang VS, Nagwani M, Holtzapple, MT. 1998. Lime pretreatment of crop residues bagasse and wheat straw. Appl Biochem Biotechnol 74: 131-159.

Chang VS, Nagwani M, Kim CH, Holtzapple MT. 2001. Oxidative lime pretreatment of high-lignin biomass. Appl Biochem Biotechnol 94:1-28.

Chen H, Ferrari C, Angiuli M, Yao J, Raspi C, Bramanti E. 2010. Qualitative and quantitative analysis of wood samples by Fourier transform infrared spectroscopy and multivariate analysis. Carbohydr Polym 82:772-778.

Chong IG, Jun CH. 2005. Performance of some variable selection methods when multicollinearity is present. Chemometrics Intelligent Lab Syst 78:103-112.

Chundawat SPS. 2009. Ph.D Thesis: Ultrastructural and Physiochemical Modifications within Ammonia Treated Lignocellulosic Cell Walls and their Influence on Enzymatic Digestibility. Lansing, MI, USA: Michigan State University.

Chundawat SPS, Venkatesh B, Dale BE. 2007. Effect of particle size based separation of milled corn stover on AFEX pretreatment and enzymatic hydrolysis. Biotechnol Bioeng 96:219-231.

Collier WE, Shultz TP, Kalasinsky VF. 1992. Infrared study of lignin: Reexamination of aryl-alkyl ether C-O stretching peak assignment. Holzforschung 46:523-528.

Colthup NB, Daly LH, Wiberley SE. 1964. Introduction to infrared and raman spectroscopy. New York, New York: Academic Press.

Cooley JW, Tukey OW. 1965. An algorithm for the machine calculation of complex Fourier series. Math Comput 19:297-301.

Corgié SC, Smith HM, Walker LP. 2011. Enzymatic transformations of cellulose assessed by quantitative high-throughput Fourier transform infrared spectroscopy (QHT-FTIR). Biotechnol Bioeng 108:1509-1520.

Dale BE, Moreira J. 1982 Freeze-explosion technique for increasing cellulose hydrolysis. Biotech Bioeng Symp. 12: 31-43.

Davin LB, Patten AM, Jourdes M, Lewis NG. 2008. Lignins: A twenty-first century challenge. In: Himmel ME, editor. Biomass recalcitrance. Deconstructing the plant cell wall for bioenergy. Oxford, United Kingdom: Blackwell Publishing Ltd. p 213-305.

de Jong S. 1993. SIMPLS: An alternative approach to partial least squares regression. Chemometrics Intelligent Lab Syst 18:251-263.

de O. Buanafina MM. 2009. Feruloylation in grasses: Current and future perspectives. Mol Plant 2:861-872.

Decker SR, Siika-aho M, Viikari L. 2008. Enzymatic depolymerization of plant cell wall hemicelluloses. In: Himmel ME, editor. Biomass Recalcitrance. Deconstructing the plant cell wall for bioenergy. Oxford, United Kingdom: Blackwell Publishing Ltd. p 352-373.

Ebringerova A. 2006. Structural diversity and application potential of hemicelluloses. Macromol Symp 232:1-12.

Efron B, Tibshirani R. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1:54-75.

Eklund R, Galbe M, Zacchi G. 1995. The influence of SO_2 and H_2SO_4 impregnation of willow prior to steam pretreatment. Bioresour Technol 52:225-229.

Eriksson L, Johansson E, Kettaneh-Wold N, Wold S. 2001. Multi- and Mega Variate Data Analysis. Principles and Applications. Umea, Sweden: Umetrics AB.

Eriksson T, Börjesson J, Tjerneld F. 2002. Mechanism of surfactant effect in enzymatic hydrolysis of lignocellulose. Enzyme Microb Technol 31:353-364.

Fan LT, Lee YH, Beardmore DR. 1981. The influence of major structural features of cellulose on rate of enzymatic hydrolysis. Biotechnol Bioeng 23:419-424.

Fan LT, Lee YH, Gharpuray MM. 1982. The nature of lignocellulosics and their pretreatments for enzymatic hydrolysis. Adv Biochem Eng 23:157-187.

Faust TD, Ibsen KN, Dayton DC, Hess JR, Kenney KE. 2008. The Biorefinery. In: Himmel ME, editor. Biomass recalcitrance. Deconstructing the plant cell wall for bioenergy. Oxford, United Kingdom: Blackwell Publishing Ltd. p 7-37.

Ferraz A, Baeza J, Rodriguez J, Freer J. 2000. Estimating the chemical composition of biodegraded pine and eucalyptus wood by DRIFT spectroscopy and multivariate analysis. Bioresour Technol 74:201-212.

Geladi P, Kowalski BR. 1986. Partial least-squares regression: a tutorial. Anal Chim Acta 185:1-17.

Ghosh A, Ghosh BK, Trimino-Vazquez H, Eveleigh DE, Montenecourt BS. 1984. Cellulase secretion from a hypercellulolytic mutant of *Trichderma reesei* RUT-30. Arch Microbiol 140:126-133.

Gilbert HJ, Stålbrand H, Brumer H. 2008. How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. Curr Opin Plant Biol 11:1-11.

Gollapalli LE, Dale BE, Rivers DM. 2002. Predicting digestibility of ammonia fiber explosion (AFEX)-treated rice straw. Appl Biochem Biotechnol 98-100:23-35.

Grabber JH. 2005. How do lignin composition, structure, and cross-linking affect degradability? A review of cell wall model studies. Crop Sci 45:820-831.

Grethlein HE. 1985. The effect of pore-size distribution on the rate of enzymatic-hydrolysis of cellulosic substrates. Bio-Technology 3:155-160.

Grethlein, HE, Converse A. 1981. Partial acid hydrolysis of poplar wood as a pretreatment for enzymatic hydrolysis. Biotech Bioeng Symp 11:67-77.

Grohmann K, Torget R, Himmel ME. 1986. Optimization of dilute acid pretreatment of biomass. Biotech Bioeng Symp 15 (Symp Biotechnol Fuels Chem, 7th, 1985): 59-80.

Guo GL, Hsu, DC, Chen, WH, Chen, WH, Hwang, WS. 2009. Characterization of enzymatic saccharification for acid-pretreated lignocellulosic materials with different lignin composition. Enz Microbial Technol 45:80-87.

Harris PJ, Stone BA. 2008. Chemistry and molecular organization of plant cell walls. In: Himmel ME, editor. Biomass recalcitrance. Deconstructing the plant cell wall for bioenergy. Oxford, United Kingdom: Blackwell Publishing Ltd. p 61-93.

Hatfield R, Vermerris W. 2001. Lignin formation in plants. The dilemma of linkage specificity. Plant Physiol 126:1351-1357.

Himmel ME, Picataggio SK. 2008. Our challenge to acquire deeper understanding of biomass recalcitrance and conversion. In: Himmel ME, editor. Biomass recalcitrance. Deconstructing the plant cell wall for bioenergy. Oxford, United Kingdom: Blackwell Publishing Ltd. p 1-6.

Höskuldsson A. 1988. PLS regression methods. J Chemometrics 2:211-228.

Höskuldsson A. 1995. A combined theory for PCA and PLS. J Chemometrics 9:91-123.

Hsu TA. 1996. Pretreatment of biomass. In: Wyman CE, editor. Handbook on Bioethanol: Production and Utilization. Briston, PA: Taylor and Francis. p 179-212.

Johnson DK, Elander R. 2008. Pretreatments for enhanced digestibility of feedstocks. In: Himmel ME, editor. Biomass Recalcitrance. Deconstructing the plant cell wall for bioenergy. Oxford, United Kingdom: Blackwell Publishing Ltd. p 435-453.

Jolliffe IT. 2002. Principal component analysis. New York: Springer-Verlag.

Kamm B, Kamm M. 2004. Principles of biorefineries. Appl Microbiol Biotechnol 64:137-145.

Karr WE, Holtzapple M. 2000. Using lime pretreatment to facilitate the enzymatic hydrolysis of corn stover. Biomass Bioenergy 18:189-199.

Keller FA, Hamilton JE, Nguyen QA. 2003. Microbial pretreatment of biomass: Potential for reducing severity of thermochemical biomass pretreatment. Appl Biochem Biotechnol 105:27-41.

Kim TH, Lee YY. 2005a. Pretreatment and fractionation of corn stover by ammonia recycle percolation process. Bioresour Technol 96:2007-2013.

Kim TH, Lee YY. 2005b. Pretreatment of corn stover by soaking in aqueous ammonia. Appl Biochem Biotechnol 121:1191-1131.

Klemm D, Heublein B, Fink H-P, Bohn A. 2005. Cellulose: Fascinating biopolymer and sustainable raw material. Angew Chem Int Ed 44:3358-3393.

Kristensen JB, Thygesen LG, Felby C, Jørgensen H, Elder T. 2008. Cell-wall structural changes in wheat straw pretreated for bioethanol production. Biotechnol Biofuels 1. Available at http://www.biotechnologyforbiofuels.com/1754-6834/1/5

Kuga S, Brown RM. 1991. Physical structure of cellulose microfibrils: Implications for biogenesis. In: Cadence Haigler H, Wiemer PJ, editors. Biosynthesis and biodegradation of cellulose. New York: Marcel Dekker, Inc. p 125-142.

Kumar R, Wyman C. 2009a. Effect of xylanase supplementation of cellulase on digestion of corn stover solids prepared by leading pretreatment technologies. Bioresour Technol 100:4203-4213.

Kumar R, Wyman C. 2009b. Effects of cellulase and xylanase enzymes on the deconstruction of solids from pretreatment of poplar by leading technologies. Biotechnol Prog 25:302-314.

Laureano-Perez L, Teymouri F, Alizadeh H, Dale B. 2005. Understanding factors that limit enzymatic hydrolysis of biomass: Characterization of pretreated corn stover. Appl Biochem Biotechnol 124:1081-1099.

Mansfield SD, Mooney CA, Saddler J. 1999. Substrate and enzyme characteristics that limit cellulose hydrolysis. Biotechnol Prog 15:804-816.

Markov AV, Gusakov AV, Kondratyeva EG, Okunev ON, Bekkarevich AO, Sinitsyn AP. 2005. New effective method for analysis of component composition of enzyme complexes from *Trichoderma reesei*. Biochemistry (Moscow) 70:657-663.

Martens H, Naes T. 1989. Multivariate calibration. Chichester: John Wiley & Sons Inc.

McCann MC, Chen I, Roberts K, Kemsley EK, Sene C, Carpita NC, Stacey NJ, Wilson RH. 1997. Infrared microspectroscopy: Sampling heterogeneity in plant cell wall composition and architecture. Physiol Plant 100:729-738.

McMillan JD. 1994. Pretreatment of lignocellulosic biomass. In: Himmel ME, Baker JO, Overend RP, editors. Enzymatic Conversion of Biomass for Fuel Production. ACS Symposium Series 566. Washington, DC: American Chemical Society. p 292-324.

Merino ST, Cherry J. 2007. Progress and challenges in enzyme development for biomass utilization. Adv Biochem Eng Biotechnol 108:95-120.

Millett MA, Baker AJ, Satter LD. 1975. Pretreatments to enhance chemical, enzymatic, and microbiological attack of cellulosic materials. Biotechnol Bioeng:193-219.

Mohnen D, Bar-Peled M, Somerville C. 2008. Cell Wall Polysaccharide Synthesis. In: Himmel ME, editor. Biomass Recalcitrance. Deconstructing the plant cell wall for bioenergy. Oxford, United Kingdom: Blackwell Publishing Ltd. p 94-187.

Moros J, Garrigues S, de la Guardia M. 2010. Vibrational spectroscopy provides a green tool for multi-component analysis. Trends Anal Chem 29:578-591.

Mosier N, Wyman C, Dale B, Elander R, Lee YY, Holtzapple M, Ladisch M. 2005. Features of promising technologies for pretreatment of lignocellulosic biomass. Bioresour Technol 96:673-686.

Murnen HK, Balan V, Chundawat SPS, Bals B, Sousa LdC, Dale BE. 2007. Optimization of ammonia fiber expansion (AFEX) pretreatment and enzymatic hydrolysis of *Miscanthus x giganteus* to fermentable sugars. Biotechnol Prog 23:846-850.

Nelson M, O'Connor RT. 1964. Relation of certain infrared bands to cellulose crystallinity and crystal lattice type. Part II. A new infrared ratio for estimation of crystaillinity in celluloses I and II. J Appl Polym Sci 8:1325-1341.

Nguyen QA, Tucker MP, Keller FA, Eddy FP. 2000. Two-stage dilute-acid pretreatment of softwoods. Appl Biochem Biotechnol 84-86:561-576.

Nicolai BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron K, Lammertyn J. 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biol Technol 46:99-118.

O'Sullivan A. 1997. Cellulose: the structure slowly unravels. Cellulose 4:173-207.

Ohgren K, Bura R, Saddler J, Zacchi G. 2007. Effect of hemicellulose and lignin removal on enzymatic hydrolysis of steam pretreated corn stover. Bioresour Technol 98:2503-2510.

Ooshima H, Burns DS, Converse AO. 1990. Adsorption of cellulase from *Trichderma reesei* on cellulose and lignacious residue in wood pretreated by dilute sulfuric acid with explosive decompression. Biotechnol Bioeng 36:446-452.

Palonen H. 2004. Ph.D Thesis: Role of Lignin in the Enzymatic Hydrolysis of Lignocellulose. Helsinki, Finland: Helsinki University of Technology.

Pandey KK, Pitman AJ. 2003. FTIR studies of the changes in wood chemistry following decay by brown-rot and white-rot fungi. Int Biodeterior Biodegrad 52:151-160.

Pauly M, Keegstra K. 2008. Cell-wall carbohydrates and their modification as a resource for biofuels. The Plant Journal 54(4):559-568.

Pavlostathis SG, Gossett JM. 1985. Alkaline treatment of wheat straw for increasing anaerobic biodegradability. Biotechnol Bioeng 27:334-344.

PerkinElmer Life and Analytical Sciences. 2005. FTIR spectroscopy. Attenuated Total Reflectance (ATR). Perkin Elmer: Technical Notes. Shelton CT: http://shop.perkinelmer.com/content/TechnicalInfo/TCH_FTIRATR.pdf

Perkins WD. 1987. Fourier transform infrared spectroscopy. Part I: Instrumentation. Topics in chemical instrumentation. J Chem Educ 63:A5-A10.

Prior BA, Day DF. 2007. Hydrolysis of ammonia-pretreated sugar cane bagasse with cellulase, β-glucosidase and hemicellulase preparations. Appl Biochem Biotechnol 146:151-164.

Qiang X, Ding S-Y, Nimlos MR, Johnson DK, Himmel ME. 2005. Atomic and electronic structures of molecular crystalline cellulose Iß: A first principles investigation. Macromol 38:10580-10589.

Raiskila S, Pulkkinen M, Laakso T, Fagerstedt K, Löija M, Mahlberg R, Paajanen L, Ritschkoff A-C, Saranpää P. 2007. FTIR spectroscopic prediction of Klason and acid soluble lignin variation in Norway spruce cutting clones. Silva Fennica 41:351-371.

Ralph J. 2010. Hydroxycinnamates in lignification. Phytochem Rev 9:65-83.

Robert P, Marquis M, Barron C, Guillon F, Saulnier L. 2005. FT-IR investigations of cell wall polysaccharides from cereal gains. Arabinoxylan infrared assignment. J Agric Food Chem 53:7014-7018.

Saha BC. 2003. Hemicellulose bioconversion. J Ind Microbiol Biotechnol 30:279-291.

Schell D, Farmer J, Newman M, McMillan J. 2003. Dilute-sulfuric acid pretreatment of corn stover in pilot-scale reactor. Appl Biochem Biotechnol 105:69-85.

Sewalt VJH, Glasser WG, Beauchemin KA. 1997. Lignin impact on fiber degradation. Reversal of inhibition of enzymatic hydrolysis by chemical modification of lignin and by additives. J Agric Food Chem 45:1823-1828.

Sierra R, Granda CB, Holtzapple MT. 2009. Short-term lime pretreatment of poplar wood. Biotechnol Prog 25:323-332.

Sluiter A, Hames B, Ruiz R, Scarlata C, Sluiter J, Templeton D, Crocker D. 2008. Determination of structural carbohydrates and lignin in biomass. Golden, CO: National Renewable Energy Laboratory. NREL Laboratory Procedure.

Sørensen HR, Pedersen S, Jørgensen CT, Meyer AS. 2007. Enzymatic hydrolysis of wheat arabinoxylan by a recombinant "minimal" enzyme cocktail containing β-xylosidase and novel endo-1,4-β-xylanase and α-L-arabinofuranosidase activities. Biotechnol Prog 23:100-107.

Sørensen HR, Pedersen S, Viksø-Nielsen A, Meyer AS. 2005. Efficiencies of designed enzyme combinations in releasing arabinose and xylose from wheat arabinoxylan in an industrial ethanol fermentation residue. Enz Microbial Technol 36:773-784.

Stone M. 1974. Cross validatory choice and assessment of statistical predictions. J Stat Soc B 36:111-147.

Sugiyama J, Persson J, Chanzy H. 1991. Combined infrared and electron diffraction study of the polymorphism of native celluloses. Macromol 24:2461-2466.

Sun Y, Cheng J. 2002. Hydrolysis of lignocellulosic materials for ethanol production: A review. Bioresour Technol 83:1-11.

Tamaki Y, Mazza G. 2011a. Rapid determination of carbohydrates, ash, and extractives contents of straw using attenuated total reflectance Fourier transform mid-infrared spectroscopy. J Agric Food Chem 59:6346-6352.

Tamaki Y, Mazza G. 2011b. Rapid determination of lignin content of straw using attenuated total reflectance Fourier transform mid-infrared spectroscopy. J Agric Food Chem 59:504-512.

Taniguchi M, Suzuki H, Watanabe D, Sakai K, Hoshino K, Tanaka K. 2005. Evaluation of pretreatment with *Pleurotus ostreatus* for enzymatic hydrolysis of rice straw. J Biosci Bioeng 100:637-643.

Tassinari T, Macy C, Spano L. 1980. Energy requirements and process design considerations in compression-milling pretreatment of cellulosic wastes for enzymatic hydrolysis. Biotechnol Bioeng 22:1689-1705.

Teymouri F, Laureano-Perez L, Alizadeh H, Dale B. 2005. Optimization of ammonia fiber explosion (AFEX) treatment parameters for enzymatic hydrolysis of corn stover. Bioresour Technol 96:2014-2024.

Tucker MP, Nguyen QA, Eddy FP, Kadam KL, Gedvilas LM, Webb JD. 2001. Fourier transform infrared quantitative analysis of sugars and lignin in pretreated softwood solid residues. Appl Biochem Biotechnol 91-93:51-61.

Van Soest P, Robertson J, Lewis B. 1991. Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. J Dairy Sci 74:3583-3597.

Wehrens R. 2011. Chemometrics with R. Multivariate data analysis in the natural sciences and life sciences. Gentleman R, Hornik K, Parmigiani G, editors. Berlin: Springer-Verlag Berlin and Heidelberg.

Williams DH, Fleming I. 1980. Spectroscopic methods in organic chemistry. New York, NY: McGraw Hill.

Wilson DB. 2008. Aerobic microbial cellulase systems. In: Himmel ME, editor. Biomass recalcitrance. Deconstructing the plant cell wall for bioenergy. Oxford, United Kingdom: Blackwell Publishing Ltd. p 374-386.

Wold H. 1966. Estimation of principal components and related models by iterative least squares. In: Krishnaiaah P, editor. Multivariate Analysis. New York: Academic Press. p 391-420.

Wold S, Sjostrom M, Eriksson L. 2001. PLS-regression: a basic tool of chemometrics. Chemometrics Intelligent Lab Syst 58:109-130.

Wyman CE, Dale B, Elander R, Holtzapple M, Ladisch M, Lee Y. 2005a. Coordinated development of leading biomass pretreatment technologies. Bioresour Technol 96:1959-1966.

Wyman CE, Dale B, Elander R, Holtzapple M, Ladisch M, Lee Y, Mitchison C, Saddler J. 2009. Comparative sugar recovery and fermentation data following pretreatment of poplar wood by leading technologies. Biotechnol Prog 25:333-339.

Wyman CE, Dale BE, Elander RT, Holtzapple M, Ladisch MR, Lee YY. 2005b. Comparative sugar recovery data from laboratory scale application of leading pretreatment technologies to corn stover. Bioresour Technol 96:2026-2032.

Xu J, Cheng J, Sharma-Shivappa RR, Burns JC. 2010. Sodium hydroxide pretreatment of switchgrass for ethanol production. Energ Fuel 24:2113-2119.

Yang B, Wyman CE. 2004. Effect of xylan and lignin removal by batch and flowthrough pretreatment on the enzymatic digestibility of corn stover cellulose. Biotechnol Bioeng 86:88-98.

Zhao Y, Wang Y, Zhu JY, Ragauskas A, Deng Y. 2008. Enhanced enzymatic hydrolysis of spruce by alkaline pretreatment at low temperature. Biotechnol Bioeng 99:1320-1328.

Zhong C, Lau MW, Balan V, Dale B, Yuan Y. 2009. Optimization of enzymatic hydrolysis and ethanol fermentation from AFEX-treated rice straw. Appl Microbiol Biotechnol 84:667-676.

Zhou G, Taylor G, Polle A. 2011. FTIR-ATR-based prediction and modeling of lignin and energy contents reveals independent intra-specific variation of these traits in bioenergy poplars. Plant Methods 7(9). Available at http://www.plantmethods.com/content/7/1/9.

Zhu, L. 2005. Ph.D Thesis: Fundamental Study of Structural Features Affecting Enzymatic Hydrolysis of Lignocellulosic Biomass. College Station, TX, USA. Texas A&M University.

CHAPTER 3

ASSESSMENT OF COMMERCIAL HEMICELLULASES FOR SACCHARIFICATION OF ALKALINE PRETREATED PERENNIAL BIOMASS[1]

Abstract

The objective of this research was to measure the effects of different cellulase and hemicellulase mixtures on fermentable sugar production from two different perennial biomasses — switchgrass and a low-impact, high-diversity prairie biomass mixture (LIHD). Each was subjected to NaOH pretreatment, followed by hydrolysis with a commercial cellulase and β-glucosidase mixture [CB] supplemented with either of two hemicellulases. For both biomasses, there was little gain in sugar yield when using CB alone beyond 20–25 mg/g TS; further gain in yield was possible only through hemicellulase supplementation. An equation that modeled CB and hemicellulase effects as occurring independently fit the data reasonably well, except at the lowest of cellulase loadings with hemicellulase, where synergistic interactions were evident. Examination of the marginal effectiveness of enzyme loadings (incremental grams sugar per incremental mg enzyme) over a broad range of loadings suggests that there is no need to customize enzymatic hydrolysis for NaOH-pretreated switchgrass and LIHD.

Keywords: Alkaline pretreatment; Enzymatic hydrolysis; Cellulase; Hemicellulase; Lignocellulose

---

[1]This work has been published as Sills DL, Gossett JM. 2011. Assessment of commercial hemicellulases for saccharification of alkaline pretreated perennial biomass. Bioresour Technol 102:1389-1398.

3.1 Introduction

Renewable bio-ethanol produced from lignocellulosic biomass has the potential to provide the USA with up to 50% of its transportation fuel, with the benefit of increased sustainability and less green-house gas production compared to traditional fossil fuels and corn-based ethanol (Biomass Research and Development Board 2008). The most economically challenging conversion step is the transformation of carbohydrate-rich plant-cell walls into fermentable sugars (Jørgensen et al. 2007), and there is a need to integrate pretreatment and enzymatic hydrolysis processes for more efficient and less expensive sugar production (Himmel et al. 2007).

Potential plant-biomass feedstocks for bio-ethanol production include dedicated energy crops such as perennial grasses and woody biomass. This study focuses on two perennial biomasses: switchgrass and a low-impact, high-diversity mixture of prairie biomasses (LIHD). Switchgrass, a perennial grass, has been identified as a potential lignocellulosic feedstock for bio-ethanol production due to its high yields, ability to grow throughout most of the United States, carbon sequestration ability, efficient nutrient utilization, and reduced input requirements relative to annual crops such as corn (McLaughlin and Kszos 2005). However, the cultivation of polycultures, such as mixtures of native prairie biomasses, results in higher yields and increased environmental benefits compared to monocultures such as switchgrass (Hill 2007). Tilman et al. (2006) reported that an 18-species LIHD that included grasses, forbs, legumes, and woody types yielded 238% more above-ground biomass than did switchgrass when grown on degraded land. Mixed species are also more resistant to disease, require less agrochemical inputs, and are in general more stable than single species (Hill 2007). No studies to date have compared LIHD to a

monoculture perennial feedstock such as switchgrass in terms of fermentable sugar production from combined pretreatment and enzymatic hydrolysis with different enzyme mixtures.

We chose, as a representative alkaline treatment, a high-NaOH [20 g NaOH/100 g total solids (TS)], ambient-temperature process, that includes reuse of unconsumed alkali. This pretreatment was previously employed in the context of improving anaerobic production of methane from wheat straw (Pavlostathis and Gossett 1985). The authors characterized the effect of recovering and reusing unconsumed NaOH, and they determined that recycled alkali did not reduce the effectiveness of the treatment, while significantly reducing chemical input. Through careful water management and use of recycle, the process requirement for fresh NaOH approaches the consumptive demand (principally from alkali saponification of esters in lignocellulose), regardless of the alkali level (g NaOH/100 g TS) employed. Low-temperature NaOH (Xu et al. 2010; Zhao et al. 2008) and lime pretreatments (Kim and Holtzapple 2005) have been successfully applied to a number of biomass feedstocks and resulted in high fermentable sugar yields. Pretreatments operated at low temperatures consume less energy compared to high-temperature pretreatments, which may contribute to life-cycle sustainability. Low-temperature treatment also minimizes cellulose dehydration, which can adversely impact saccharification (Atalla et al. 2008).

Several researchers have recently reported that adding Multifect Xylanase® (MX) to cellulase- and β-glucosidase-mediated hydrolysis of corn stover, poplar, switchgrass, miscanthus, and rice-straw that underwent alkaline or acidic pretreatments resulted in increased glucose production (Berlin et al. 2007; Bura et al. 2009; Kumar and Wyman 2009a; Kumar and Wyman 2009b; Murnen et al. 2007; Ohgren et al. 2007; Zhong et al. 2009). Acidic treatments hydrolyze a significant portion of hemicellulosic polymers. Thus it is surprising that the addition of

hemicellulases to cellulase- and β-glucosidase-mediated hydrolysis of acid-pretreated biomasses would result in increased glucose yields, but researchers have demonstrated this for a number of substrates (Berlin et al. 2007; Bura et al. 2009; Kumar and Wyman 2009a; Kumar and Wyman 2009b; Ohgren et al. 2007). Alkaline pretreatments produce solids that retain the majority of hemicellulose polymers intact, and as expected, hemicellulase supplementation increased 72-h glucose yields for numerous substrates (Kumar and Wyman 2009a; Kumar and Wyman 2009b; Murnen et al. 2007; Ohgren et al. 2007; Prior and Day 2007; Zhong et al. 2009). Addition of Multifect Pectinase® (MP) to cellulase and β-glucosidase-mediated hydrolysis also resulted in increased glucose yields from enzymatic hydrolysis of acidic ([Berlin et al. 2007; Dien et al. 2008]) and alkaline pretreated substrates (Dien et al. 2008; Murnen et al. 2007).

MX and MP were chosen for this study based on their different relative amounts of hemicellulolytic activities. MX has higher xylanase activity than MP, while MP exhibits side-group cleaving activities (e.g., α-arabanifuranosidase and ferulic-acid esterase) as well as β-xylosidase activity that MX lacks (Dien et al. 2008). The mechanism by which hemicellulase enzymes assist cellulases is not completely understood (Gilbert et al. 2008; Sørensen et al. 2007). Researchers (Berlin et al. 2007; Dien et al. 2008) have ruled out cellulase activity in MX and MP being the cause of increased glucose production, and the current working hypothesis is that hemicellulase enzymes hydrolyze non-cellulosic carbohydrate polymers and open up the lignocellulosic matrix, thus making cellulose more accessible to cellulases (Murashima et al. 2003).

Researchers have proposed that due to the overwhelming diversity of lignocellulosic biomass, pretreatment processes and subsequent enzyme requirements will need to be tailored to different biomass substrates individually (Chandra et al. 2007; Eggeman and Elander 2005;

Merino and Cherry 2007; Wyman et al. 2005]). This has the potential to create cumbersome logistics for a prospective technology that is expected to involve the application of varying, seasonal energy crops to a single, lignocellulose-to-ethanol processing facility. While researchers have shown that feedstocks with large differences (e.g., corn stover versus poplar) require different conditions for high sugar yields (Kumar and Wyman 2009a; Kumar and Wyman 2009b), there is not enough data in the literature to say definitively that different biomasses always require different pretreatment and hydrolysis conditions. The present study focuses on the following question: to what extent must hydrolysis conditions be customized for switchgrass versus LIHD — two substrates that have the potential to be major feedstocks for cellulosic biofuel production? We addressed this question with a study that measured the effect of adding two commercial hemicellulases to cellulase- and β-glucosidase-mediated saccharification of alkaline (NaOH) pretreated switchgrass and LIHD. Additionally, we present a novel modeling approach and the use of "marginal effectiveness" as a way of looking at enzyme loadings.

3.2 Methods

3.2.1 Raw Material

Switchgrass was harvested in the spring of 2005 in Ligonier, PA, and was provided by Tom Stickle. The biomass was removed from a bale and was milled in a Thomas Wiley mill (Thomas Scientific; Swedesboro, NJ, USA) to pass through ¼-in. mesh screen. The LIHD (already milled to pass a ¼-in. mesh screen) was a provided by the Cedar Creek Ecosystem Science Reserve, which is operated by the College of Biological Sciences at the University of Minnesota. LIHD, a mixture of 38 species (listed in Table A.1 in Appendix 1), was cultivated as

described in Tilman et al. (2006), and its composition is similar to the 18-species mixture described in the same paper.

3.2.2 Chemicals and Enzymes

Glucose and xylose were from Sigma–Aldrich, 98+ % purity. Spezyme CP® cellulase (Lot 4900857805; 59 FPU/mL, 123 mg/mL), Multifect Xylanase (MX, Lot 4900876030; 42 mg protein/mL), and MultifectPectinase FE (MP, Lot 4010833580; 83 mg protein/mL) were provided by Genencor Danisco® US Inc. (Rochester, NY, USA). Novozyme® 188 β-glucosidase (Lot 01K0735; 280 CBU/mL, 140 mg protein/mL) was from Sigma–Aldrich (St. Louis, MO, USA). Protein was measured in 96-well microplates using a Sigma bicinchoninic acid (BCA) kit following manufacturer's instructions (Sigma) with bovine-serum-albumin as a standard. Cellulase activity was measured with a filter paper assay (Adney and Baker 2008).

3.2.3 Compositional Analysis

Solid samples (both before and after pretreatment) were analyzed in triplicate for glucan, xylan, and Klason-lignin content using a modification of the standard NREL LAP protocol (http://www.nrel.gov/biomass/analytical_procedures.html). Briefly, 3mL of 72% H_2SO_4 was added to 300 mg of dry biomass solids. Slurries were incubated for 1 h at 30 °C and stirred manually every 5–10 min. Water (84 mL) was added to each sample resulting in an acid concentration of 4% (w/w), and samples were autoclaved at 121 °C for 1 h. Samples were passed through glass microfiber filter disks (Whatman Grade 934-AH) placed in Gooch crucibles with medium-porosity, fritted-glass bottoms and washed five times with hot distilled water. Aliquots of the acid hydrolysates were saved for HPLC analysis of carbohydrates (see below).

Washed sample-solids were dried at 105 °C for 12 h and hot-weighed according to the method outlined in (Van Soest et al. 1991). Dried samples were ashed in a muffle furnace at 550 °C for 5 h and hot-weighed a second time. Lignin mass was calculated by subtracting the second weight (ash + crucible) from the first weight (ash + Klason lignin + crucible).

Acid hydrolysate samples were neutralized with calcium carbonate powder and filtered through a 0.2-μm nylon filter before measurement of sugars. Glucose and xylose were measured by high performance liquid chromatography (HPLC) with a Bio-Rad HPX-87P column held isothermally at 84 °C. A refractive-index detector was used with pure water as eluent. Glucose and xylose concentrations were calculated from calibration curves, which related LC peak areas to mg/L of glucose and xylose in each sample. Soluble glucose and xylose were converted to equivalent glucan and xylan units, using 0.9 g glucan/g glucose and 0.88 g xylan/g xylose.

3.2.4 Alkaline Pretreatment Conditions

Pretreatments of both substrates were conducted at alkali levels of 5, 10, and 20 g NaOH/(100 g TS), and pretreatment of switchgrass was also conducted at 50 g NaOH/(100 g TS). Pretreatments were conducted with 5% (w/w) total solids concentration at 25 °C in batch reactors on a rotary shaker at 200 rpm for 24 h. Untreated controls were prepared in batch reactors with water only, but otherwise employed the same conditions as for the alkaline treatment. When pretreatment was finished, the slurries were neutralized with 1 M solutions of HCl. Solids were collected by filtering neutralized slurries through Miracloth (Calbiochem, San Diego, CA) and washed with ten pretreatment volumes of water.

Pretreatment reactions were conducted in triplicate, and sugars released during pretreatment were measured (see Section 3.2.5) to verify pretreatment reproducibility prior to

using pretreated solids for subsequent enzymatic hydrolysis reactions. Three aliquots of pretreated solids intended for enzymatic hydrolysis reactions were analyzed for TS by drying at 105 °C for 12 h followed by 2 h in a desiccator. Additional triplicate pretreatment reactions were conducted for the purpose of measuring the amount of solids solubilized during pretreatment. TS remaining after pretreatment were measured as described above.

3.2.5 Characterization of Pretreatment Liquors

After solids separation (see Section 3.2.4), pretreatment liquors were analyzed for glucose and xylose with HPLC as described in Section 3.2.3. Sugars present in alkaline pretreatment liquors were primarily in the form of oligomers, and prior to HPLC analysis these samples underwent acid hydrolysis according to NREL standard protocol (Sluiter et al. 2008). Sulfuric acid was added to samples to create a 4% (w/w) solution and samples were autoclaved at 121 °C for 1 h. Resulting liquid hydrolysates were neutralized with calcium carbonate powder and analyzed with HPLC as described in Section 3.2.3.

Consumption of alkali during pretreatment was quantified by titrating pretreatment slurries with 0.25 M HCl to pH 8.4. Slurries were stirred for 1 h during the titration process to facilitate release of alkali, and HCl was added as necessary to achieve a stable pH of 8.4. Alkali consumption was measured in triplicate.

3.2.6 Enzymatic Hydrolysis

Enzymatic hydrolysis was conducted on the pretreated solids with cellulase and hemicellulase mixtures and conditions with a modification of the standard NREL LAP protocol (http://www.nrel.gov/biomass/analytical_procedures.html). Hydrolysis was performed with a

solids concentration of 2.5% TS (w/w) in 0.05 M citrate buffer (pH 4.8) with tetracycline (30 mg/L) and cyclohexamide (20 mg/L) to prevent microbial growth, at 50 °C, in batch reactors on a rotary shaker set at 130 rpm. A mixture of cellulase (Spezyme CP) + β-glucosidase (Novozyme188) which is denoted as CB (for "cellulase and β-glucosidase") was prepared from Spezyme CP and Novozyme188 at a protein ratio of 1:0.4 (or 1 FPU:1.75 CBU) in order to prevent end-product inhibition from possible accumulation of cellobiose. MX loadings of 0, 2, 4, 12.5, and 25 mg/g TS were added to hydrolysis reactions with six loadings of CB (7, 14, 21, 70, 140, and 210 mg/g TS, corresponding to approximately 5 FPU + 8 CBU; 10 FPU + 16 CBU; 15 FPU + 25 CBU; 50 FPU + 80 CBU; 100 FPU + 160 CBU; and 150 FPU + 250 CBU per g glucan). MP loadings of 0, 5, 10, and 30 mg/g TS were added to reactions run with two CB concentrations (7 and 21 mg/g TS, respectively). Substrate blanks without enzyme addition, and enzyme blanks without substrate additions were run in parallel. All sample reactions and blanks were conducted in triplicate. Glucose and xylose were measured with HPLC as described in Section 3.2.3.

3.2.7 Model Fitting

As an aid to understanding the degree to which cellulases and hemicellulases might act synergistically, we formulated a model based on the opposite possibility — i.e., that the two classes of enzymes contribute independently to sugar yields. The extent to which experimental data are well-fit (or not) by such a model can be suggestive of the degree to which cellulases and hemicellulases act independently of each other. The model attempts to predict glucose and xylose conversions (as percent of total glucose and xylose contents of untreated biomass) from enzymatic hydrolysis of alkaline-treated switchgrass or LIHD as a function of CB and MX

loadings. The model considers the effects of CB and MX to be completely independent; thus, there are separate, saturation-type terms for the contribution of each enzyme mixture:

$$\text{Glucose Conversion (\% of potential)} = \frac{a_G \cdot CB}{b_G + CB} + \frac{c_G \cdot MX}{d_G + MX} \qquad (3.1)$$

$$\text{Xylose Conversion (\% of potential)} = \frac{a_X \cdot CB}{b_X + CB} + \frac{c_X \cdot MX}{d_X + MX} \qquad (3.2)$$

where *a*, *b*, *c*, and *d* are model fitting parameters; and, CB and MX loadings are in mg/g TS. The parameters in the numerator (*a* and *c*) represent the upper limits to conversion (as contributed by the actions of CB and MX, respectively). The parameters in the denominator (*b* and *d*) represent the sensitivities of conversions to CB and MX at lower loadings — i.e., the CB and MX loadings, respectively, that produce one-half their maximum effects. While the model is empirical, single-term versions of this same form are used widely to model phenomena exhibiting saturation maxima — e.g., Monod kinetics and Langmuir adsorption isotherms. To our knowledge, no one has previously used a two-term saturation-type expression to model independent effects of plant cell-wall degrading enzymes.

Curve-fitting and numerical calculations of derivatives were conducted with the Surface Fitting Toolbox, Matlab (2009, Mathworks, Natick, MA). Glucose and xylose data from 72-h hydrolysis of alkaline-treated switchgrass and LIHD were each fit to Eqs. (3.1) and (3.2), respectively.

3.3 Results and Discussion

3.3.1 Compositional Changes from Pretreatment

Chemical composition (glucose, xylose, and lignin) for untreated and alkaline-pretreated (20 g NaOH/100 g TS, 25 °C, 24 h) biomasses are summarized in Table 3.1. Since arabinose, galactose and mannose make up only a small portion of the sugars in switchgrass and LIHD, only glucose and xylose were considered in this study. Raw switchgrass contained 17% and 27% more glucan and xylan, respectively, compared to raw LIHD, and 21% less Klason lignin. Alkaline treatment solubilized 6% and 13% of glucan in switchgrass and LIHD, respectively; 18% and 27% of the xylan in raw switchgrass and LIHD, respectively; and about 50% of the Klason lignin in both biomasses. These results are in agreement with previous research with wheat straw and pine (Pavlostathis and Gossett 1985; Zhao et al. 2008).

The measured glucan and xylan recovered in pretreatment liquors are also shown in Table 3.1 (solubles, as percent of original TS). Mass-balance closures (remaining particulates + measured solubles, compared to originally measured totals in raw biomasses) for glucan were 97% and 95% for switchgrass and LIHD, respectively, and mass-balance closures for xylan were 103% and 97% for switchgrass and LIHD.

Alkali consumption was 15% higher with LIHD than with switchgrass. Since previous studies with similar high-NaOH treatments found correlation between treatment effectiveness and alkali consumption (Pavlostathis and Gossett 1985), these data suggest that alkaline treatment might be more effective on LIHD than on switchgrass.

Table 3.1 Untreated and pretreated[a] switchgrass and LIHD composition, % of initial dry matter (TS) in raw biomass.

Biomass type	TS[b]	Glucan[b]		Xylan[b]		Klason Lignin[b]	Alkali Consumption[b]
		Particulate	Soluble[c]	Particulate	Soluble[c]	Particulate	(g NaOH/100 g TS)
Switchgrass (raw)	100	36.2 ± 0.2	na	20.1 ± 0.4	na	18.9 ± 1.0	na
PT Switchgrass	69.7 ± 1.2	33.9 ± 1.8	1.13 ± 0.04	16.4 ± 1.6	4.36 ± 0.0	9.81 ± 1.2	5.62 ± 0.18
LIHD (raw)	100	30.9 ± 0.2	na	15.8 ± 0.0	na	23.9 ± 0.1	na
PT LIHD	65.3 ± 1.7	26.9 ± 1.7	2.46 ± 0.05	11.6 ± 1.7	3.78 ± 0.0	11.8 ± 1.8	6.47 ± 0.14

[a] 20 g NaOH per 100 g TS
[b] Mean values of three replicates ± standard deviation
[c] Soluble glucose and xylose converted to equivalent glucan and xylan units, using 0.9 g glucan/g glucose and 0.88 g xylan/g xylose.

3.3.2 Effect of NaOH Loading on Enzymatic Hydrolysis

Our study assessed effects of pretreatment primarily in terms of glucose and xylose productions during subsequent enzymatic hydrolysis. Percent conversions from enzymatic hydrolysis that are reported in Fig. 3.1, Fig. 3.2, Fig. 3.3, Fig. 3.4, and Fig. 3.5 do not include the glucan and xylan solubilized by pretreatment. Additional fermentable sugar contributions from pretreatments are included later in this paper (Section 3.3.5), where total sugar conversions from combined pretreatment and enzymatic hydrolysis are discussed.

The effects of NaOH loading on glucose release during enzymatic hydrolysis of alkaline-treated switchgrass and LIHD solids with a cellulase/β-glucosidase-only (CB) mixture (15 FPU/25 CBU per g glucan) are shown in Fig. 3.1. Glucose and xylose conversions here are defined as the percentages of total glucan and xylan in the original biomass that appear as glucose and xylose through hydrolysis. Appropriate multipliers were used in unit conversion: 0.9 g glucan/g glucose and 0.88 g xylan/g xylose.

Glucose and xylose productions increased with NaOH loading for both switchgrass and LHID up to loadings of 20 g NaOH per 100 g of TS. The major difference between the two-biomass types was that the rates of glucose production were faster with LIHD than with switchgrass for all alkali levels. This can be seen upon examination of the first time points (4 and 6 h for LIHD and switchgrass, respectively) as well as the 24-h time points for both biomasses. Glucose conversions were at least as high at 4 h for LIHD as at 6 h for switchgrass for all alkaline-treatment levels.

After 24h of hydrolysis, glucose conversions were higher for LIHD than for switchgrass, again for all alkaline-treatment levels. Hydrolysis of LIHD solids that were pretreated with 5

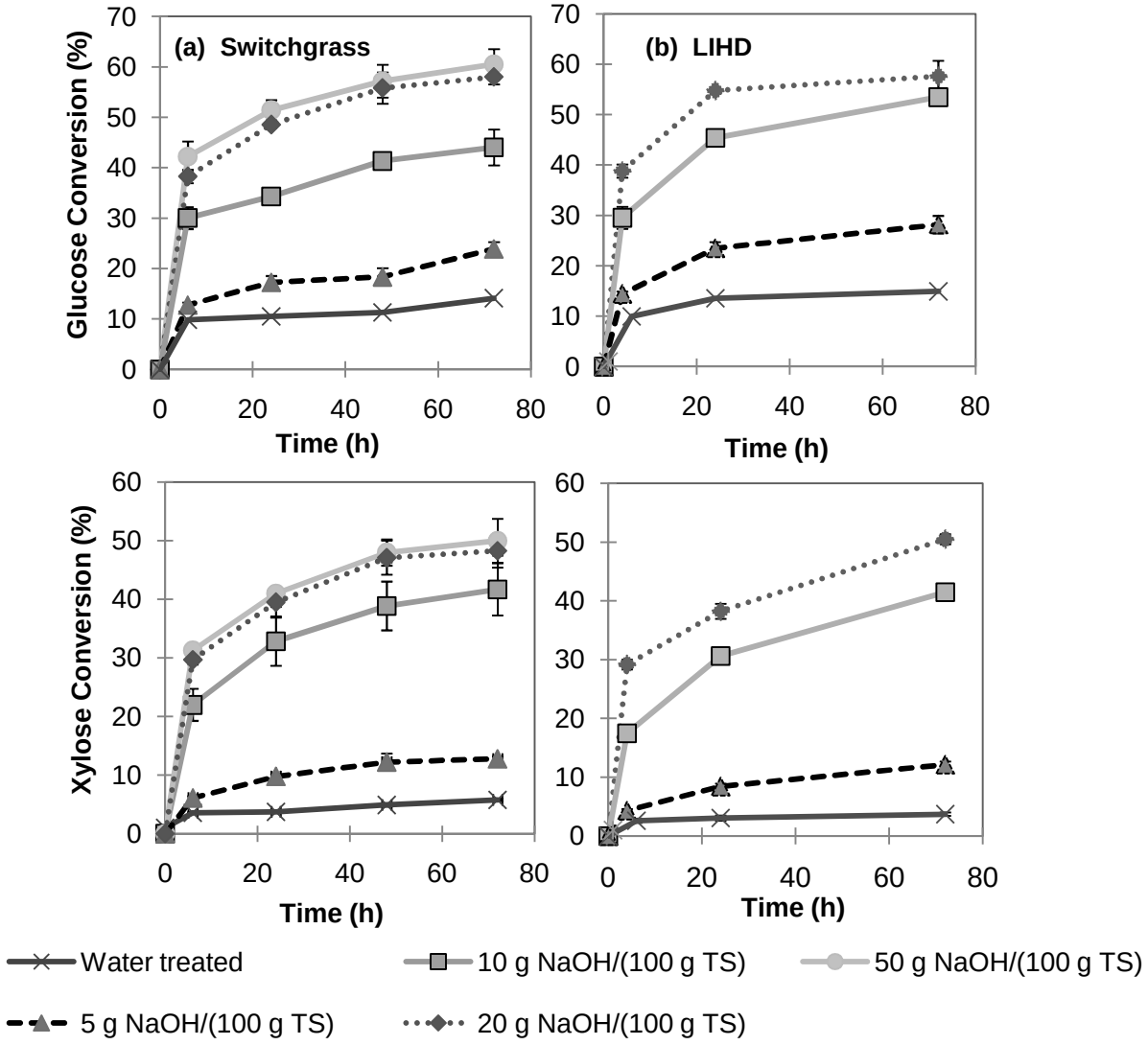

Figure 3.1 The effect of NaOH loading on glucose and xylose conversions during cellulase- and ß-glucosidase-mediated hydrolysis (21 mg/g TS or 15 FPU and 25 CBU per g glucan) of (a) switchgrass (0, 5, 10, 20, and 50 g NaOH per 100 g TS) and (b) LIHD (0, 5, 10, and 20 g NaOH per 100 g TS). Values are means of three replicates, and error bars represent standard deviations.

and 10 g NaOH per 100 g TS resulted in higher 72-h glucose conversions in subsequent enzymatic hydrolysis compared to switchgrass. On one hand, the higher glucose conversions for LIHD are expected due to the higher alkali consumption values described in the previous section. On the other hand, these results are surprising, because LIHD has 25% more Klason-lignin than does switchgrass (see Table 3.1), and lignin levels usually correlate negatively with rates and extents of digestibility of lignocellulosic biomass (Grabber 2005).

Hydrolysis of solids pretreated with 20 g NaOH per 100 g TS resulted in an approximately four-fold increase in 72-h glucose conversions from both biomass types compared to untreated biomasses. Upon examination of Fig. 3.1, one might conclude that alkaline pretreatment of LIHD should be operated with an alkali loading of 10 g NaOH per 100 g TS, but when alkali-recycle is employed, NaOH consumption approaches that of the consumptive demand from reactions occurring during pretreatment. For LIHD, the alkali consumptive demand was only 13–16% higher at a treatment level of 20 g NaOH than at 10 g NaOH per 100 g TS (data not shown). Thus, we chose 20 g NaOH per 100 g TS for subsequent experiments with hemicellulase supplementation because it gave the highest conversions with both substrates, and there was no substantial increase in glucose and xylose conversions above a treatment level of 20 g NaOH per 100 g TS.

3.3.3 Hemicellulase Supplementation

The effects of CB (7, 14, and 21mg/g TS), MX (0, 2, 4, 12.5, and 25 mg/g TS), and MP (0, 5, 10, 30 mg/g TS) loadings on 4-h glucose and xylose conversions (as a percent of the glucose and xylose in the raw biomasses) from hydrolysis of NaOH-pretreated switchgrass and LIHD are presented in Fig. 3.2. Glucose conversions at 4 h were consistently higher for LIHD,

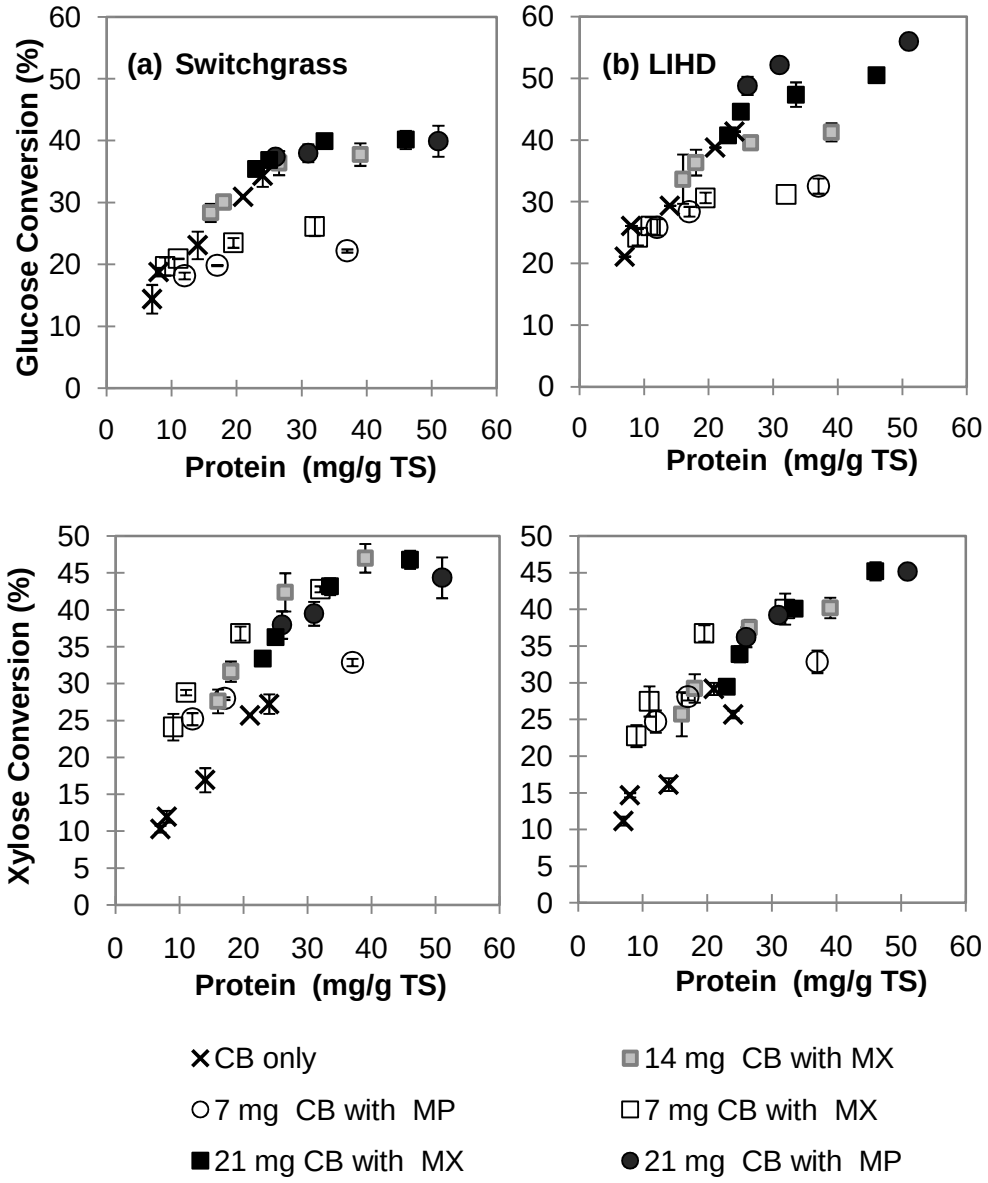

Figure 3.2 Effect of CB (7, 14, and 21 mg/g TS), MX (0-25 mg/g TS), and MP (0-30 mg/g TS) loadings on 4-h glucose and xylose conversions from enzymatic hydrolysis of NaOH pretreated (20 g NaOH per 100 g TS, 24 h, 25 ºC) (a) switchgrass and (b) LIHD. Cellulase to ß-glucosidase ratio is 1 FPU:1.75 CBU. Values are means of three replicates, and error bars represent standard deviations.

with conversions of 20–55%, compared to switchgrass, with conversions of 15–40% — with or without MX or MP supplementation, suggesting that the cellulose in LIHD is more accessible and/or digestible than the cellulose in switchgrass. At the lowest CB loading of 7 mg/g TS, 4-h glucose conversions from switchgrass and LIHD were improved more by providing additional CB than by supplementing with MX or MP. At high CB loadings, effectiveness of hemicellulase addition differed between substrates: at 21 mg CB/g TS, hemicellulase supplementation gave negligible improvement in 4-h glucose conversions from switchgrass, while for LIHD, MX and MP supplementation each resulted in increases from a base case of 40% up to 55% of the theoretical glucose.

There were fewer differences between biomasses in their 4-h xylose-conversion responses to increased loadings of CB versus MX or MP supplementation (see Fig. 3.2). Either MX or MP supplementation resulted in greater increases in xylose production than did increased CB loading alone, as one might expect since MX and MP have greater activity on xylans than does CB (Dien et al. 2008; and as determined in the present study, data not shown).

Glucose conversions from 72-h enzymatic hydrolysis of NaOH-pretreated switchgrass and LIHD with CB only, CB + MX, and CB + MP are presented in Fig. 3.3. Note that two charts are presented for glucose conversion with each substrate (the uppermost chart highlighting the low-enzyme-loading region for clarity). Along with the data are curves based on a hypothetical model that assumes independent actions of CB and MX (Eq. (3.1)). Note that the model was fit simultaneously, not sequentially, to the CB-only and CB + MX data — i.e., the model parameters were determined that best fit all of the data. The solid black curves represent model predictions for CB-only, and the colored curves branching off the solid black curves represent predicted additional conversions from MX supplementation. The CB + MX

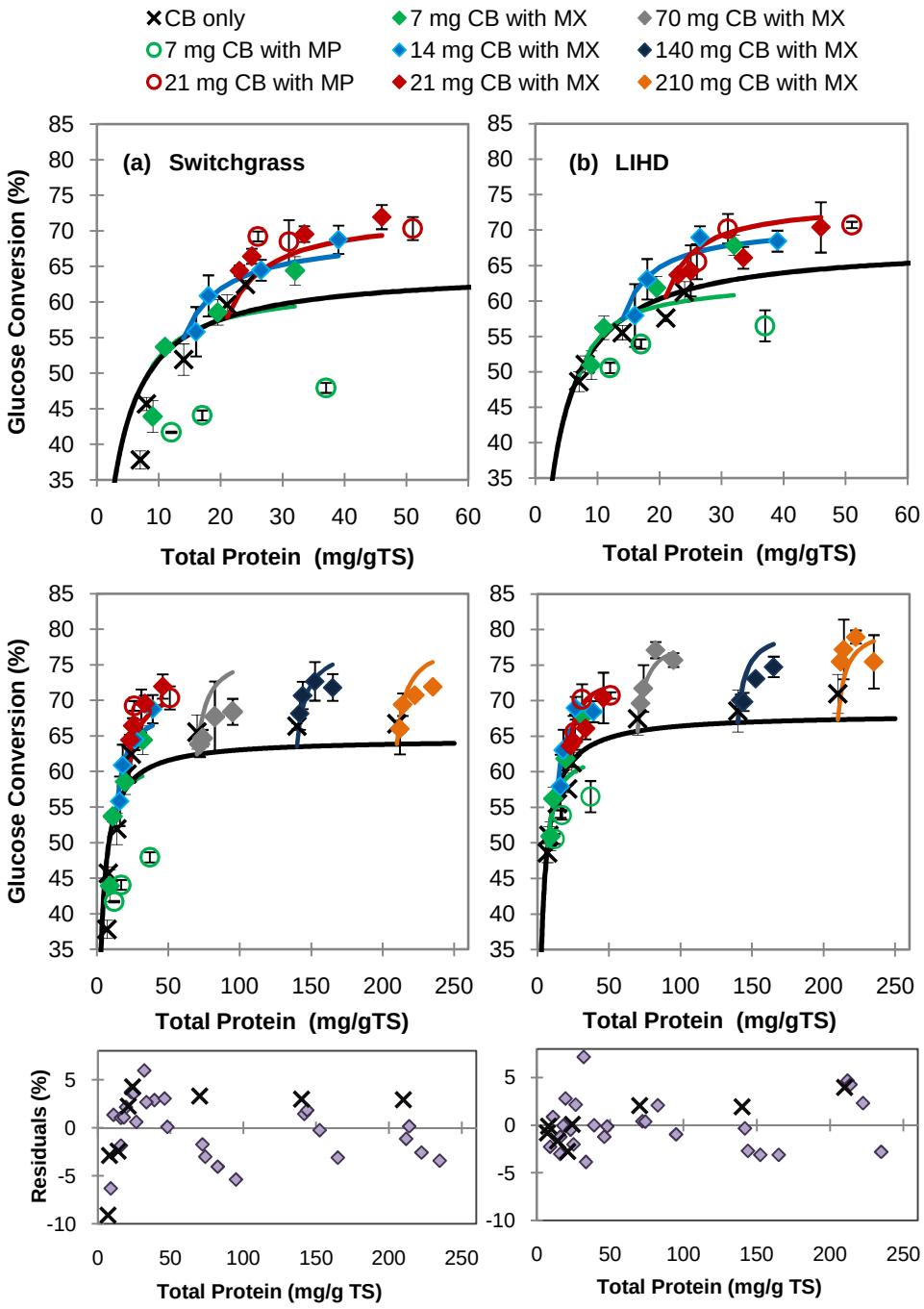

Figure 3.3 Effect of CB (7, 14, 21, 70, 140, and 210 mg/g TS) with and without MX (2, 4, 12.5, and 25 mg/g TS) and MP (5, 10, and 30 mg/gTS) on 72-h glucose conversions from hydrolysis of NaOH–pretreated (20 g per 100 g TS, 24 h, 25˚C) (a) switchgrass and (b) LIHD. Model fits for CB-only and CB + MX data presented as lines in same colors as data points with matching CB loadings. Cellulase to β-glucosidase ratio is 1 FPU:1.75 CBU. Data values are means of three replicates, and error bars represent standard deviations. Residuals (vertical distance difference between data values and model-predicted values) presented below data with CB-only residuals as black Xs and CB + MX residuals as closed diamonds.

predictions are presented in the same colors as the experimental data from reactions with the matching CB loading (7, 14, 21, 70, 140, and 210 mg/g TS). CB + MP data were not modeled, since it is apparent that MP was significantly less effective than MX at lower CB loadings, and no better than MX at higher CB loadings.

Eq. (3.1) was fit to the CB-only and CB + MX data for both biomass types in part to help explain differences between biomasses and enzyme mixtures. Parameter values ± 95% confidence intervals for switchgrass and LIHD, respectively, are as follows: a_G = 64.6 ± 2.9 and 68.2 ± 2.3; b_G = 2.64 ± 0.66 and 2.62 ± 0.47; c_G = 14.6 ± 5.8 and 13.46 ± 3.9; and d_G = 6.75 ± 8.7 and 5.49 ± 5.4. Visual inspection of the model curves, the data, and residual plots suggest that besides the case of the CB + MX data at the lowest CB loading (7 mg/g TS), the model curves fit the data for both biomasses fairly well, with a better fit for LIHD. The goodness-of-fit parameters confirm this observation, with root mean square error (RMSE) values of 3.5 and 2.7 and R^2 values 0.93 and 0.97 for switchgrass and LIHD, respectively.

Eq. (3.1) under-predicts glucose conversions from CB + MX at CB loadings of 7 mg/g TS for both biomasses, suggesting that the effects of CB and MX are not independent as is assumed by the model. In other words, MX was more effective at increasing glucose conversions at the lowest CB loadings compared to all other CB loadings. This increased effectiveness of MX at low CB loadings agrees with previous research that reported increased synergism between and cellulases and hemicellulases at lower enzyme loadings (Selig et al. 2008b). A closer inspection of the CB-only data for switchgrass reveals that the model over-predicts glucose conversions at lower CB-only loadings (7, 8, and 14 mg/g TS), while under-predicting glucose conversions at higher CB-only loadings (70, 140, and 210 mg/g TS). This also suggests that for switchgrass, MX supplementation is having a greater benefit at lower CB

loadings than at higher CB loadings, skewing the parameters of the CB term in Eq. (3.1) to better fit over-all an independent-effect model. A logical interpretation is that the relatively small contribution of hemicellulases in the CB mixture (Dien et al. 2008; and as determined in the present study, data not shown) contributes significantly to hemicellulase activity at higher CB loadings, making the effect of hemicellulase supplementation less than is observed at low CB loading. Thus, the effects of MX and CB are not strictly independent over the entire range examined, though the two-term, independent-effects model does a reasonable job over-all of fitting the data, as is evident from the lack of trend in residuals with respect to the full set of data over the entire range of protein-loadings.

In general for both substrates, at the lowest CB loading of 7 mg/g TS, adding more CB or supplementing with MX resulted in similar increases in glucose conversions, suggesting that at these loadings, protein is simply protein. However, beyond total protein loadings of about 20–25 mg/g TS, the CB + MX data rose above the CB-only curve. This means that adding supplemental MX resulted in a greater incremental conversion than would have been achieved by adding the same mg of additional CB beyond 20–25 mg/g TS. Protein is not simply protein: MX is enhancing glucose conversion in some manner apart from what would have been achieved by simply adding more CB. This observation agrees with results presented by Berlin et al. (2007) who demonstrated that for dilute-acid-pretreated corn stover at high enzyme loadings, supplementing cellulase with MX and MP reduced enzyme requirements (compared to reactions with cellulase alone) required for high glucose conversions, but at low enzyme loadings glucose conversions were not affected by type of protein (cellulase or hemicellulase).

CB-only data and models for both substrates indicate that there was little benefit, in terms of 72-h glucose conversions, to raising CB loadings above 20–25 mg/g TS. Substantial

improvements beyond this protein loading were realized only through MX supplementation, with loadings between 12.5 and 25 mg MX/g TS appearing optimal. The plateau observed in CB effectiveness is in agreement with other researchers, who observed virtually no differences in glucose and xylose conversions between cellulose + β-glucosidase loadings of 15 + 25 and 60 + 120 (FPU + CBU) per g glucan (approximately equal to 21 and 81 mg/g TS loading of CB in the present study) (Wyman et al. 2005).

In an attempt to better fit the CB-only data, we tried an alternate, sequential, two-step, model-fitting scheme: we fit the first term of Eq. (3.1) to CB-only data, thus determining the values of a_G and b_G. Parameters a_G and b_G were then kept constant, and CB-only and CB + MX data were fit to both terms of Eq. (3.1), thus determining values for c_G and d_G (model not shown). As expected because of the greater constraints in fitting, this sequential procedure resulted in a model with an over-all higher sum of square error (for both biomasses), and the model curve did not follow the CB + MX data very well — especially for switchgrass. However, the two-step sequential-fit model described the CB-only data better than the model presented in Fig. 3.3, again suggesting that CB and MX effects are not completely independent, especially for switchgrass. We chose to use the independent-effects model of Eq. (3.1), as fit in Fig. 3.3, for further analyses presented in Section 3.3.4. It is likely accurate enough for practical use at all but the lowest of CB loadings.

Differences between biomasses in terms of 72-h glucose conversions were most pronounced at the lowest enzyme loadings. At CB-only loadings of 7 and 8 mg/g TS, glucose conversions were higher for LIHD compared to switchgrass, however conversions increased more with addition of more CB or hemicellulase supplementation for switchgrass than LIHD, resulting in similar yields for both biomasses at protein loadings higher than 14 mg/g TS. At the

highest CB loadings (70, 140, and 210 mg/g TS), MX supplementation resulted only in small increases in 72-h glucose conversions for switchgrass with maximum conversions of approximately 70%, suggesting that for switchgrass there were no benefits in terms of 72-h glucose conversions from raising CB loadings above 21 mg/g TS, and combined CB + MX loadings above 50 mg/g TS. However, MX supplementation of hydrolysis reactions of LIHD with CB loadings of 70, 140, and 210 mg/g TS, resulted in increases in 72-h glucose conversions with maximum conversions of approximately 80% compared to 70% from reactions with MX added to a CB loading of 21 mg/g TS.

Xylose conversions from 72-h hydrolysis of NaOH-pretreated switchgrass and LIHD with CB only, CB + MX, and CB + MP are presented in Fig. 3.4 along with model curves, based on Eq. (3.2), that were fit simultaneously to the CB-only and CB + MX data. As in the case of glucose, CB + MP data were not fit to model curves. Parameter values ± 95% confidence intervals for switchgrass and LIHD, respectively, are as follows: a_X = 51.6 ± 2.9 and 50.9 ± 2.5; b_X = 2.31 ± 0.76 and 3.09 ± 0.73; c_X = 15.7 ± 3.9 and 14.6 ± 3.2; and d_X = 2.49 ± 2.4 and 2.79 ± 2.2. Visual inspection of the model curves, the data, and residual plots suggest that the model curves fit the data for both biomasses fairly well. The goodness-of-fit parameters confirm this observation, with RMSE values of 3.4 and 2.9 and R^2 values 0.93 and 0.94 for switchgrass and LIHD, respectively. As in the case of 4-h sugar conversions presented in Fig. 3.2, MX or MP supplementation was superior to adding more CB in terms of their xylose-conversion responses. Furthermore, differences between biomasses in terms of their 72-h xylose-conversion responses to adding additional CB or supplementing with MX or MP were less significant than differences in their 72-h glucose conversion responses.

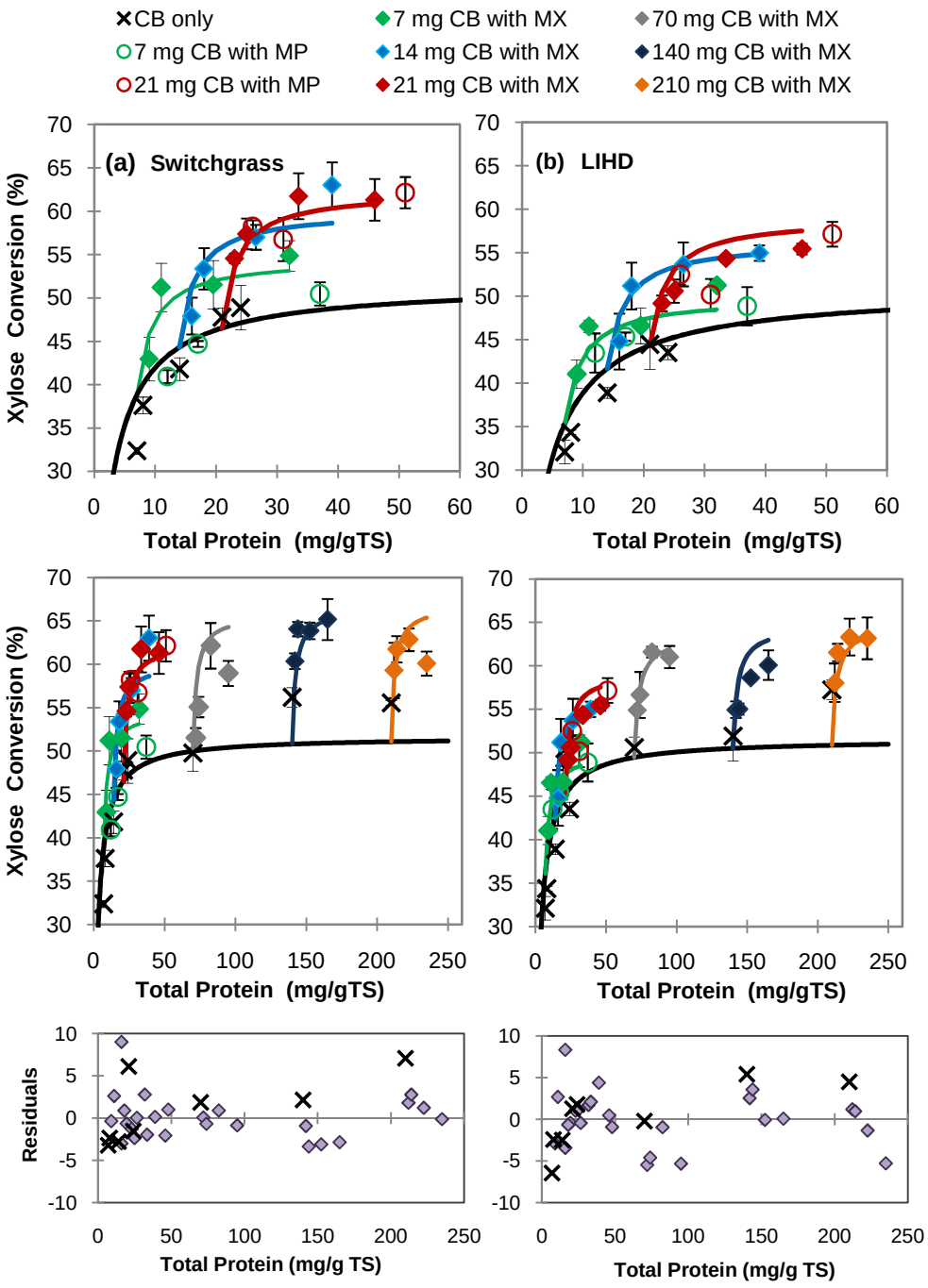

Figure 3.4 Effect of CB (7, 14, 21, 70, 140, and 210 mg/g TS) with and without MX (2, 4, 12.5, and 25 mg/g TS) and MP (5, 10, and 30 mg/gTS) on 72-h xylose conversions from hydrolysis of NaOH–pretreated (20 g per 100 g TS, 24 h, 25˚C) (a) switchgrass and (b) LIHD. Model fits for CB-only and CB + MX data presented as lines in same colors as data points with matching CB loadings. Cellulase to β-glucosidase ratio is 1 FPU:1.75 CBU. Data values are means of three replicates, and error bars represent standard deviations. Residuals are presented below data with CB-only residuals as black Xs and CB + MX residuals as closed diamonds.

The model equations (based on Eq. (3.2)) for 72-h xylose conversions fit the CB + MX data fairly well at CB loadings of 7, 14, and 21 mg/g TS (see Fig. 3.4). However, as in the case of glucose, the model under-predicted xylose conversions at CB-only loadings of 70, 140, and 210 mg/g TS, but to a greater extent than for glucose. Once again, a logical interpretation is that the relatively small concentration of hemicellulases in the CB mixture contributes significantly to hemicellulase activity at higher CB loadings. Furthermore, since hemicellulase directly releases xylose, it increased concentration in higher-CB reactions should have a greater effect on xylose conversions than on glucose conversions, which explains the larger deviations between xylose data and model predictions compared to those for glucose (see Fig. 3.3 and Fig. 3.4). And so, as in the case of glucose, the effects of MX and CB on 72-h xylose conversions are not strictly independent, though the two-term, independent-effects model fits the data reasonably well over the full range of loadings studied.

3.3.4 Optimal Enzyme Loadings

Do NaOH-pretreated switchgrass and LIHD require customized enzyme mixtures and loadings for optimal process performance? With respect to hemicellulase supplementation, MX was generally superior to MP with both biomass substrates. While there were some conditions in our study where the two hemicellulases performed similarly, in many other cases, MX clearly out-performed MP. MX loadings in the range between 12.5 and 25 mg/g TS are recommended for both switchgrass and LIHD, with the higher end of the range suggested for use at low CB loading. MX loadings higher than 25 mg/g TS were not evaluated, based on previous research that showed negative effects on glucose and xylose conversions with higher MX loadings (Berlin et al. 2007; Bura et al. 2009).

How, then, might the two substrates differ in terms of optimal CB loading? The optimal enzyme loading occurs where the cost of adding more enzyme per incremental amount of sugar produced equals the over-all unit cost of sugar production. Adding more enzyme is only cost-effective if doing so lowers the over-all unit cost of sugar production. Besides a dependence on enzyme cost and effectiveness, other fixed and operating costs matter also. An important consideration in determining optimal enzyme loading in any specific process context would be knowledge of the *marginal effectiveness*, defined as the incremental grams of sugar produced per incremental gram of enzyme applied. This is not an economic study, and we did not identify fixed and operating costs associated with pretreatment and hydrolysis, but we can use the model described in Sections 3.2.7 and 3.3.3 to measure differences in marginal sugar production versus incremental increases in protein loading for NaOH-pretreated switchgrass and LIHD. Previous researchers have suggested that for industrial-scale cellulosic ethanol to be economically viable, both hexose and pentose sugars will need to be utilized (Merino and Cherry 2007; Wyman 2003). Thus we chose to look at marginal *total* sugar production (glucose + xylose). And because MX loadings at all CB loadings exhibited a diminishing return beyond 12.5 mg MX/g TS, we chose a constant MX loading of 12.5 mg/g TS for this analysis.

Fig. 3.5(a) presents 72-h glucose + xylose yields from hydrolysis reactions at a fixed MX loading of 12.5 mg/g TS with different CB loadings (7, 14, 21, 70, 140, and 210 mg/g TS). The curves in Fig. 3.5(a) depicts model fits of total sugars using the parameters determined through fitting glucose and xylose data previously (Fig. 3.3 and Fig. 3.4). Note that sugar yields presented in Fig. 3.5(a) are expressed as percent of TS of the original dry raw biomass (rather than in % of total potential sugar in raw biomass, as was done in Fig. 3.1, Fig. 3.2, Fig. 3.3, and

Fig. 3.4). This change in units facilitates the determination of marginal effectiveness, as described below.

First, note that total sugar yields for switchgrass are everywhere higher than for LIHD, when expressed on a per-TS basis. This is largely due to the higher sugar content of switchgrass compared to LIHD, and the similar percent conversions (in units of % potential glucose and xylose) for both biomasses presented in Fig. 3.3 and Fig. 3.4 and discussed in Section 3.3.3. We constructed similar plots for reactions run with alternative MX loadings of 0, 2, 4, and 25 mg/g

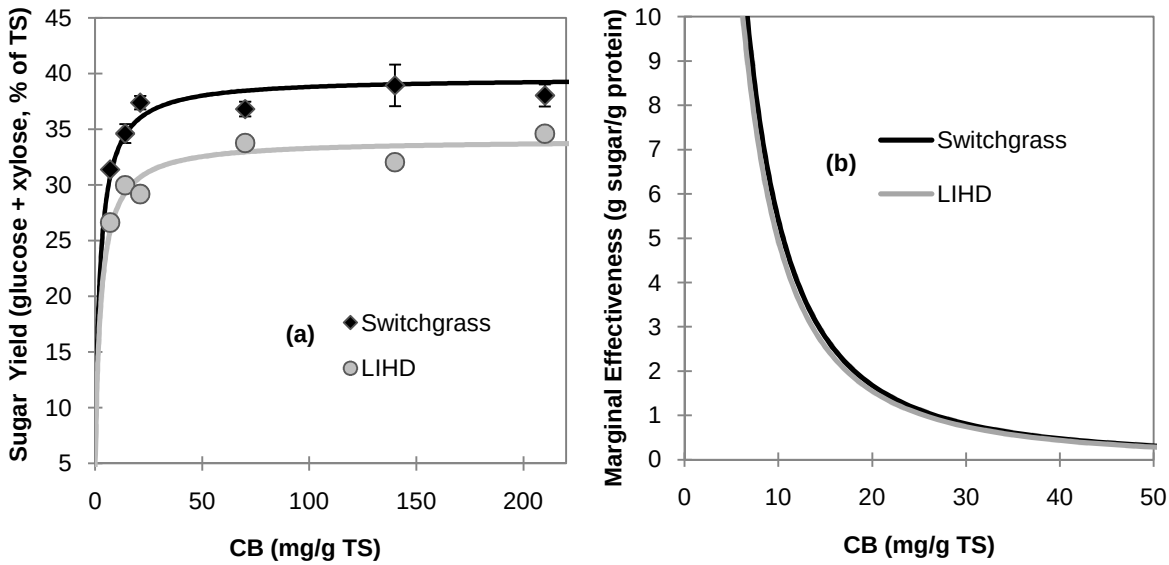

Figure 3.5 (a) Effect of CB loading (7, 14, 21, 70, 140, 210 mg/g TS) on 72-h sugar yields (glucose + xylose, % of TS) of NaOH-pretreated (20 g per 100 g TS, 24 h, 25˚C) switchgrass and LIHD at a fixed MX loading of 12.5 mg/g TS. Solid lines represent model curves based on Eq. (1) + Eq. (2). Data values are means of three replicates, and error bars represent standard deviations. (b) Marginal effectiveness of enzyme addition versus CB loading, plots constructed by calculating dervatives of respective curves in (a).

TS (data not shown), and in 29 out of a total of 30 cases, 72-h total sugar yields (as percent TS) were higher for switchgrass compared to LIHD. We used the Wilcoxon Signed Rank Test and determined that sugar yields (per g TS) from switchgrass were higher than yields from LIHD for CB only and CB + MX treatments at a significance level above 99.9% (Conover 1971).

Fig. 3.5(b) presents the marginal effectiveness of enzyme addition (g additional sugar per g additional enzyme protein) versus CB loading for both substrates, which were determined by calculating the derivatives of the curves presented in Fig. 3.5(a). Fig. 3.5(b) shows that increasing CB protein loadings during hydrolysis at a fixed MX loading of 12.5 mg/g TS results in essentially identical marginal increases in 72-h monosaccharide yields for switchgrass and LIHD. Thus, we conclude that NaOH-pretreated switchgrass and LIHD would have the same optimal enzyme loadings, if other costs of production were similar between them. Note, however, that as costs unrelated to enzyme loading go higher, the optimal enzyme loading would move to higher values. Thus, optimal loading is undoubtedly context-dependent – with the context changing rapidly as new uses/values for byproducts are considered. Our point simply is that the two substrates have similar marginal effectiveness with respect to enzyme loading, despite their having significantly different 72-h sugar yields (on a per mass dry solids basis).

3.3.5 Total Sugar Conversions

The main focus of this study was to determine the effects of hemicellulase addition on sugar conversions from enzymatic hydrolysis of switchgrass and LIHD that underwent alkaline and acidic pretreatments. However, if one wishes to ascertain the total glucan and xylan conversions, one must add the sugars released during pretreatment to those released in enzymatic

hydrolysis. For alkaline pretreatment (20 g NaOH per 100 g TS) and enzymatic hydrolysis with a CB loading of 21 mg/g TS and MX loading of 12.5 mg/g TS, fermentable-sugars conversions were as follows: 73% and 74% of the glucan of switchgrass and LIHD, respectively; and 84% and 78% of the xylan of switchgrass and LIHD, respectively (see Fig. 3.3, Fig. 3.4, and Table 3.1) were liberated from combined pretreatment and enzymatic hydrolysis processes. Most of the carbohydrates released during NaOH pretreatment were oligomers, but we verified that these were hydrolysable with the cellulase and hemicellulase mixtures used in this study (data not shown).

3.4 Conclusion

In hydrolysis of NaOH-pretreated switchgrass and LIHD, hemicellulase supplementation was required to achieve maximum sugar conversions. An equation that modeled CB and MX effects independently fit the data reasonably well, except at the lowest cellulase loadings, where synergistic interactions were evident. The data and model indicated that marginal enzyme effectiveness (incremental sugar produced per incremental enzyme added) was the same for switchgrass and LIHD over a wide range of CB-loadings. Thus, our research suggests that, although absolute sugar yields differed between the biomasses, it is not necessary to customize hemicellulase supplementation and cellulase loadings for the two substrates studied here.

Acknowledgements

This study was supported with funding from the Department of Transportation and the Northeast Sun Grant Initiative. We thank Larry Walker and the Biofuels Research Laboratory at Cornell University for providing analytical equipment and facilities; James Robertson and

Michael Van Amburgh in the Department of Animal Science at Cornell University for their facilities and instruction in biomass characterization; Tom Stickle of Ligonier, PA for providing switchgrass; Jason Hill from the Institute of the Environment at the University of Minnesota, and Matt Ziehr from the Cedar Creek Ecosystem Science Reserve at the University of Minnesota for the LIHD; and Genencor Danisco for enzyme mixtures.

REFERENCES

Adney B, Baker J. 2008. Measurement of cellulase activities. Golden, CO: National Renewable Energy Laboratory. NREL Laboratory Analytical Procedure.

Atalla RH, Brady JW, Matthews JF, Ding S-Y, Himmel ME. 2008. Structure of plant cell wall celluloses. In: Himmel ME, editor. Biomass recalcitrance: deconstructing the plant cell wall for bioenergy. Oxford, United Kingdom: Blackwell Publisher Ltd. p 188-212.

Berlin A, Maximenko V, Gilkes N, Saddler J. 2007. Optimization of enzyme complexes for lignocellulosic hydrolysis. Biotechnol Bioeng 77:287-295.

Biomass Research and Development Board, 2008. National Biofuels Action Plan. <http://www1.eere.energy.gov/biomass/pdfs/nbap.pdf>.

Bura R, Chandra RP, Saddler J. 2009. Influence of xylan on the enzymatic hydrolysis of steam-pretreated corn stover and hybrid poplar. Biotechnol Prog 25:314-322.

Chandra RP, Bura R, Mabee WE, Berlin A, Pan X, Saddler JN. 2007. Substrate pretreatment: the key to effective enzymatic hydrolysis of lignocellulosics? Adv Biochem Eng Biotechnol 108:67-93.

Conover WJ. 1971. Practical nonparametric statistics. New York: John Wiley and Sons.

Dien B, Ximenes EA, O'Bryan PJ, Moniruzzaman M, Li X, Balan V, Dale B, Cotta MA. 2008. Enzyme characterization of AFEX and liquid hot-water pretreated distillers' grains and their conversion to ethanol. Bioresour Technol 99:5216-5225.

Eggeman T, Elander RT. 2005. Process and economic analysis of pretreatment technologies. Bioresour Technol 96:2019-2025.

Gilbert HJ, Stålbrand H, Brumer H. 2008. How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. Curr Opin Plant Biol 11:1-11.

Grabber JH. 2005. How do lignin composition, structure, and cross-linking affect degradability? A review of cell wall model studies. Crop Sci. 45:820-831.

Hill J. 2007. Environmental costs and benefits of transportation biofuel production from food- and lignocellulosic-based energy crops. A review. Agron Sustain Dev 27:1-12.

Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. 2007. Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science 315:804-807.

Jørgensen H, Kristensen JB, Felby C. 2007. Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. Biofuels, Bioprod Bior 1:119-134.

Kim S, Holtzapple MT. 2005. Lime pretreatment and enzymatic hydrolysis of corn stover. Bioresour Technol 96:1994-2006.

Kumar R, Wyman C. 2009a. Effect of xylanase supplementation of cellulase on digestion of corn stover solds prepared by leading pretreatment technologies. Bioresour Technol 100:4203-4213.

Kumar R, Wyman C. 2009b. Effects of cellulase and xylanase enzymes on the deconstruction of solids from pretreatment of poplar by leading technologies. Biotechnol Prog 25:302-314.

McLaughlin SB, Kszos LA. 2005. Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. Biomass Bioenerg 28:515-535.

Merino ST, Cherry J. 2007. Progress and challenges in enzyme development for biomass utilization. Adv Biochem Eng Biotechnol 108:95-120.

Murashima K, Kosugi A, Doi RH. 2003. Synergistic effects of cellulosomal xylanase and cellulases from Clostridium cellulovorans on plant cell wall degradation. J Bacteriol 185:1518-1524.

Murnen HK, Balan V, Chundawat SPS, Bals B, Sousa LdC, Dale BE. 2007. Optimization of ammonia fiber expansion (AFEX) pretreatment and enzymatic hydrolysis of *Miscanthus x giganteus* to fermentable sugars. Biotechnol Prog 23:846-850.

Ohgren K, Bura R, Saddler J, Zacchi G. 2007. Effect of hemicellulose and lignin removal on enzymatic hydrolysis of steam pretreated corn stover. Bioresour Technol 98:2503-2510.

Pavlostathis SG, Gossett JM. 1985. Alkaline treatment of wheat straw for increasing anaerobic biodegradability. Biotechnol Bioeng 27:334-344.

Prior BA, Day DF. 2007. Hydrolysis of ammonia-pretreated sugar cane bagasse with cellulase, β-glucosidase and hemicellulase preparations. Appl Biochem Biotechnol 146:151-164.

Selig MJ, Knoshaug EP, Adney WS, Himmel ME, Decker SR. 2008. Synergistic enhancement of cellobiohydrolase performance on pretreated corn stover by addition of xylanase and esterase activities. Bioresour Technol 99:4997-5005.

Sørensen HR, Pedersen S, Jørgensen CT, Meyer AS. 2007. Enzymatic hydrolysis of wheat arabinoxylan by a recombinant "minimal" enzyme cocktail containing β-xylosidase and novel endo-1,4-β-xylanase and α-L-arabinofuranosidase activities. Biotechnol. Prog. 23:100-107.

Tilman D, Hill J, Lehman C. 2006. Carbon negative biofuels from low-input high-diversity grassland biomass. Science 314:1598-1600.

Van Soest P, Robertson J, Lewis B. 1991. Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. J Dairy Sci 74:3583-3597.

Wyman C. 2003. Potential synergies and challenges in refining cellulosic biomass to fuels, chemicals, and power. Biotechnol. Prog. 19:254-262.

Wyman CE, Dale BE, Elander RT, Holtzapple M, Ladisch MR, Lee YY. 2005. Comparative sugar recovery data from laboratory scale application of leading pretreatment technologies to corn stover. Bioresour Technol 96:2026-2032.

Xu J, Cheng J, Sharma-Shivappa RR, Burns JC. 2010. Sodium hydroxide pretreatment of switchgrass for ethanol production. Energ Fuel 24:2113-2119.

Zhao Y, Wang Y, Zhu JY, Ragauskas A, Deng Y. 2008. Enhanced Enzymatic Hydrolysis of Spruce by Alkaline Pretreatment at Low Temperature. Biotechnol Bioeng 99:1320-1328.

Zhong C, Lau MW, Balan V, Dale B, Yuan Y. 2009. Optimization of enzymatic hydrolysis and ethanol fermentation from AFEX-treated rice straw. Appl Microbiol Biotechnol 84:667-676.

CHAPTER 4

USING FTIR TO PREDICT SACCHARIFICATION FROM ENZYMATIC HYDROLYSIS OF ALKALI-PRETREATED BIOMASSES

Abstract

Fourier transform infrared, attenuated total reflectance (FTIR-ATR) spectroscopy combined with partial least squares (PLS) regression accurately predicted 72-h glucose and xylose conversions (g sugars/100 g potential sugars) and yields (g sugars/100 g dry solids) from cellulase-mediated hydrolysis of alkali-pretreated lignocellulose. Six plant biomasses that represent a variety of potential biofuel feedstocks — two switchgrass cultivars, big bluestem grass, a low-impact, high-diversity mixture of 32 species of prairie biomasses, mixed hardwood, and corn stover— were subjected to four levels of low-temperature NaOH pretreatment to produce 24 samples with a wide range of potential digestibility. PLS models were constructed by correlating FTIR spectra of pretreated samples to measured values of gluose and xylose conversions and yields. Variable selection, based on 90% confidence intervals of regression-coefficient matrices, improved the predictive ability of the models, while simplifying them considerably. Final models predicted sugar conversions with coefficient of determination for cross-validation (Q^2) values of 0.90 for glucose and 0.89 for xylose, and sugar yields with Q^2 values of 0.92 for glucose and 0.91 for xylose. The sugar-yield models are noteworthy for their ability to predict enzymatic saccharification per mass dry solids without *a priori* knowledge of the composition of the solids. All peaks retained in the final regression coefficient matrices were previously assigned to chemical bonds and functional groups in lignocellulose, demonstrating that the models were based on real chemical information. This study demonstrates that FTIR

spectroscopy combined with PLS regression can be used to rapidly estimate sugar conversions and yields from enzymatic hydrolysis of pretreated plant biomass.

Keywords: lignocellulose, FTIR, PLS regression

4.1 Introduction

Replacing fossil-based liquid fuels with biofuels made from renewable plant materials (e.g. switchgrass and woody materials) will lower greenhouse-gas emissions (Farrell et al. 2006) and stimulate rural economies (Somerville 2006), while reducing our dependence on foreign oil. The transformation of plant-biomass feedstocks into fermentable sugars is the largest technical hurdle preventing the development of a full-scale, cellulosic-biofuel industry (Jørgensen et al. 2007; Wyman 2007). To make cellulosic ethanol cost-competitive, we need to develop efficient pretreatment and enzymatic hydrolysis processes that produce concentrated streams of fermentable sugars. Combined pretreatment and enzymatic hydrolysis schemes are limited by the recalcitrance of lignocellulosic plant cell walls in which crystalline cellulose, itself resistant to biodegradation, is embedded in an impenetrable matrix of hemicellulose and lignin. Cellulose is composed of long chains of glucose linked via β-(1-4) glycosidic bonds; hemicellulose is made up of branched, short-chain carbohydrate polymers composed of hexoses, pentoses, and/or uronic acids; and lignin is a high-molecular-weight, randomly linked phenolic polymer. In addition, the chemical compositions of hemicellulose and lignin polymers in plant cell walls vary depending on type of species, cultivars, and tissues, as well as location of cultivation, time of harvest, and duration of storage, making the design of effective pretreatment and enzymatic hydrolysis schemes challenging (Chandra et al. 2007; Merino and Cherry 2007).

Standard wet-chemistry analyses [e.g. NREL two-step, acid-hydrolysis protocol (Sluiter et al. 2008), and the acid-detergent method (Van Soest et al. 1991)] quantify glucan (or cellulose), xylan (or hemicellulose), and lignin. However, they are labor-intensive and do not differentiate among types of cellulose (e.g. I or II), hemicellulose (e.g. xyloglucan or arabinoxylan) or lignin (e.g. syringyl or guaicyl), and do not provide insight into chemical bonds that link these three components to one another. Researchers have demonstrated that the forms of cellulose (Atalla et al. 2008), hemicellulose (Gilbert et al. 2008; Saha 2003), lignin (Grabber 2005), and carbohydrate-lignin bonds (Casler and Jung 2006) affect recalcitrance and, in turn, sugar yields from enzymatic hydrolysis. Therefore there is need for a method that characterizes lignocellulose based upon chemical attributes associated with recalcitrance, allowing prediction of sugar production attainable from pretreated biomass without having to conduct trials of enzymatic hydrolysis — or perhaps without having to conduct wet-chemistry analyses at all.

Fourier transform infrared (FTIR) spectroscopy has been used qualitatively to characterize effects of chemical and biological processing on lignocellulose (Gwon et al. 2010; Kristensen et al. 2008; Pandey and Pitman 2003). Recently, Corgié et al. (2011) used FTIR to study enzymatic hydrolysis of pure cellulose, correlating sugar yields directly to previously assigned cellulose peaks. Unfortunately, the complexity of lignocellulose and the similarity of types of bonds and functional groups across polymer types (e.g. glycosidic bonds exist in both cellulose and hemicellulose) result in IR spectra that are difficult to interpret straightforwardly.

Chemometric techniques that include multivariate models (e.g. partial least squares [PLS] and principal components [PC] regressions) can be applied to complex and collinear data (e.g. IR spectra) to extract relevant information. Both PC and PLS regressions reduce large spectral data sets by combining collinear variables (e.g. wavenumber intensities) into a small number of latent

variables (LVs), which are then used in place of the full data set to build regression models. PC and PLS models have been constructed from IR spectra to quantify cellulose, hemicellulose and lignin in raw (Chen et al. 2010;) and pretreated (Tucker et al. 2001) wood; lignin, extractives, and ash in raw straw (Tamaki and Mazza 2011a; Tamaki and Mazza 2011b); and lignin in grasses (Allison 2009) and wood (Raiskila et al. 2007; Zhou et al. 2011). Such models were constructed by correlating FTIR spectra to the chemical constituent of interest as measured by wet-chemistry techniques.

To our knowledge, only one published study exists where multivariate regression was applied to FTIR spectra of pretreated biomass in an attempt to estimate saccharification from enzymatic hydrolysis. Gollapalli et al. (2002) found that FTIR combined with PC regression did not adequately predict glucose and xylose yields from hydrolysis of rice straw that had been pretreated by ammonia-fiber explosion (AFEX). However, it is not clear to what extent they explored alternative techniques for normalization of FTIR spectra and for removing unimportant LVs, techniques that can improve PC regression models (Raiskila et al. 2007).

We used FTIR spectroscopy, combined with PLS regression, to predict 72-h glucose and xylose productions from cellulase- and β-glucosidase-mediated hydrolysis of six lignocellulosic substrates that underwent four levels of a low-temperature, alkaline pretreatment. The six substrates, chosen to capture the variability present in potential plant biofuel feedstocks, consisted of two switchgrass cultivars; big bluestem grass; a low-impact, high-diversity mixture of prairie biomasses; mixed hardwood; and corn stover. We chose four levels of pretreatment (0, 5, 10, and 20 g NaOH per 100 g TS) to generate substrates with a wide range of digestibility.

4.2 Materials and Methods

4.2.1 Biomass Samples

Two switchgrass cultivars (SG1 and SG2) were among the six biomasses. SG1 was harvested in Ligonier, PA, in Spring 2005. SG2 was harvested near Ithaca, NY, in Fall 2009. Big bluestem grass [BBS] was harvested near Ithaca, NY, in Fall 2009. The low-impact, high-diversity mixture of 32 species of prairie biomasses (LIHD) was obtained from the Cedar Creek Ecosystem Science Reserve and was cultivated as described previously (Sills and Gossett 2011; Tilman et al. 2006). Mixed hardwood [MHW] was obtained from MESA Inc., Auburn, NY, harvested in 2007. Corn stover [CS] was obtained from the National Renewable Laboratory (NREL), Golden, CO, in 2009. The MHW, CS, and LIHD were previously milled to pass a ¼-inch mesh screen, and the remaining biomasses were milled in a Thomas Wiley mill (Thomas Scientific; Swedesboro, NJ) to pass a ¼-inch mesh screen.

4.2.2 Chemicals and Enzymes

Glucose and xylose were from Sigma-Aldrich, 98+% purity. Spezyme CP® cellulase (60 FPU/mL, 118 mg protein/mL) was provided by Genencor Danisco® US Inc. (Rochester, NY, USA). Novazyme® 188 ß-glucosidase (280 CBU/mL, 140 mg protein/mL) was from Sigma Aldrich (St. Louis, MO, USA).

4.2.3 Compositional Analysis

Glucan, xylan, and Klason lignin in solid samples (both before and after pretreatment) were quantified using a modification of the standard NREL LAP protocol (http://www.nrel.gov/biomass/analytical_procedures.html; Sills and Gossett 2011). Resulting

glucose and xylose in acid hydrolyzates were quantified via high performance liquid chromatography (HPLC) (Sills and Gossett, 2011). These were considered to be the potential glucose and xylose in the pretreated solids. For purposes of describing solids compositions, they were converted to equivalent, polymeric glucan and xylan using conversion factors of 0.9 and 0.88, respectively.

4.2.4 Pretreatment and Enzymatic Hydrolysis

Pretreatments of all biomasses were conducted at the following alkali levels: 0, 5, 10, and 20 g NaOH per 100 g TS with 5% (w/w) total solids concentration at 25 °C in batch reactors on a rotary shaker at 200 rpm for 24 hours as described previously (Sills and Gossett 2011). Solids were separated by filtering neutralized slurries through Miracloth (Calbiochem, San Diego, CA) prior to enzymatic hydrolysis.

Enzymatic hydrolysis was conducted on pretreated solids with a modification of the standard NREL LAP protocol (http://www.nrel.gov/biomass/analytical_procedures.html; Sills and Gossett 2011). Hydrolysis was performed at 50 ºC in batch reactors on a rotary shaker at 130 rpm with a solids concentration of 2.5% TS (w/w) in 0.05 M citrate buffer (pH 4.8) with tetracycline (30 mg/L) and cyclohexamide (20 mg/L) to prevent microbial growth. Triplicate reactions were run with cellulase (Spezyme CP) and β-glucosidase (Novozyme 188) at loadings of 15 and 6 mg protein per g TS, respectively, corresponding to approximately 15 FPU + 25 CBU per g glucan. Triplicate substrate blanks without enzyme addition and enzyme blanks without substrate addition were run in parallel. Glucose and xylose were measured with HPLC (Sills and Gossett, 2011). We then calculated two expressions of saccharification for each sugar: (1) glucose (or xylose) *conversion*, defined as the percentage of the potential glucose (or xylose)

solubilized in enzymatic hydrolysis; and (2) glucose (or xylose) *yield*, defined as the g glucose (or xylose) solubilized per 100 g of pretreated TS. The former quantifies the accessibility of the sugar components in the solid substrate; the latter is perhaps a more useful parameter for the practitioner.

4.2.5 FTIR-ATR Spectroscopy

Prior to FTIR analysis, samples were dried under vacuum at 50 ºC for 24 h. Triplicate spectra of pretreated solids samples were collected by attenuated total reflectance (ATR) using a Vertex 70 FTIR spectrometer (Bruker Optics, Germany) equipped with a MIRacle-ATR sampling device (Pike Technologies, Madison, WI). Samples were pressed against a single-reflection diamond crystal with a torque knob that ensured identical pressure was applied to all samples. Spectra were obtained by averaging 64 scans from 4000–400 cm^{-1} at a resolution of 2 cm^{-1}, and corrected for background absorbance by subtraction of the spectrum of the empty ATR crystal. The ATR sample chamber was purged with N_2 and held under vacuum during sampling, and background samples were taken every 15 minutes. Spectra were vector-normalized, ATR-corrected, and baseline-corrected using a concave rubber-band correction to account for penetration depth and frequency variations using the OPUS software (Bruker, Germany). Triplicate spectra were averaged and converted into text files in OPUS for further analysis.

4.2.6 Construction of Partial Least Squares Regression Models

All statistical analyses were conducted with the software package R 2.9.2 (R Core Development Team, Vienna, Austria, 2009). The average FTIR-ATR spectra in the "fingerprint" range of 800–1800 cm^{-1} were combined into a single data matrix (X) with dimensions of 24

samples by 1039 wavenumber intensities, using the *ChemoSpec* package (Hanson, B. 2010, ChemoSpec: Exploratory Chemometrics for Spectroscopy. R package version 1.45), and the values of 72-h glucose and xylose productions (as either conversions or yields) were denoted as response vectors (Y). PLS regression models were developed with the *pls* package (Mevik and Wehrens 2007) by correlating mean-centered X matrices to mean-centered Y vectors using the SIMPLS algorithm (de Jong 1993) coupled with cross-validation using the leave-one-out (LOO) procedure (Geladi and Kowalski 1986a; Geladi and Kowalski 1986b). For LOO cross-validation, each sample was removed from the data set and then predicted using the model built without it. This was done 24 times to obtain a root mean square error of cross validation (RMSECV) and a coefficient of determination for validation (Q^2). The number of LVs included in each PLS model was chosen to minimize validation errors in Y, estimated with the RMSECV. The root mean square error of calibration (RMSEC) and the coefficient of determination for calibration (R^2) were then calculated for each model.

4.2.7 Variable Selection

The aim of variable selection is to reduce the number of independent variables (i.e. wavenumber intensities) by eliminating regions of the spectra that do not explain the response variables (i.e. glucose and xylose productions). Reducing variables has been shown to improve the predictive ability of multivariate regression models while making them simpler (i.e. using a small number of LVs) and thus easier to interpret (Andersen and Bro 2010; Chong and Jun 2005; Wold et al. 2001). We attempted to select variables based on variables important for projection (VIP) scores (Chong and Jun 2005) and 90% confidence intervals (CIs) of the regression coefficients (Wehrens 2011). After testing multiple models via trial and error, we concluded

that, for our work, variable selection was best done by removing regions of the spectra with 90% CIs of regression coefficients that spanned the value of zero — regardless of VIP scores. Since there are no analytical solutions for standard errors of PLS regression coefficients, 90% (CIs) of the regression coefficients were calculated using a nonparametric bootstrapping technique with a percentile method (Efron and Tibshirani 1986; Wehrens 2011). Briefly, two thousand bootstrap samples were generated, and each was a subset of the data, where some samples from the data set were present multiple times, while others were absent, and regression coefficients were calculated for each model. The 5^{th} and 95^{th} percentiles of each regression coefficient were then used to define 90 % CIs.

4.3 Results and Discussion

4.3.1 Compositional Analysis

Glucan, xylan, and lignin contents of raw and pretreated biomasses are presented in Table 4.1. Glucan content of raw biomasses ranged from 31 to 43% of total solids for the six-biomass samples. Xylan content ranged from 15 to 24%; the four grass species (SG1, SG2, CS, and BBS) all contained over 20% xylan, while LIHD and MHW each contained approximately 15% xylan. Lignin content ranged from 14% for CS to 24% for LIHD and MHW; lignin content of SG1, SG2, and BBS were all approximately 20%. Detailed data concerning solubilization effects of pretreatment appear in Table A.2 in Appendix 2.

4.3.2 Glucose and Xylose Conversions From Enzymatic Hydrolysis

Glucose and xylose conversions from 72-h enzymatic hydrolysis of pretreated solids with cellulase and β-glucosidase are presented in Fig. 4.1. Conversions here are defined as the

Table 4.1 Glucan, xylan, and lignin contents (mean value of triplicates ± standard deviations) of raw and pretreated biomass solids, % of remaining dry matter.

Pretreatment		Switchgrass 1	Switchgrass 2	Big Bluestem	LIHD	Mixed Hardwood	Corn Stover
Raw	Glucan	36.2 ± 0.2	35.0 ± 0.3	37.9 ± 0.4	30.9 ± 0.2	42.8 ± 0.2	39.2 ± 0.7
	Xylan	20.1 ± 0.4	23.5 ± 0.1	21.1 ± 0.4	15.8 ± 0.0	14.6 ± 0.2	23.2 ± 0.7
	Lignin	18.9 ± 1.0	21.5 ± 0.1	19.9 ± 0.3	23.9 ± 0.2	23.9 ± 0.2	13.5 ± 0.7
0 g NaOH /100 g TS	Glucan	38.7 ± 0.4	37.6 ± 0.6	39.3 ± 0.7	31.8 ± 1.1	44.3 ± 0.7	41.1 ± 0.1
	Xylan	22.6 ± 0.2	24.3 ± 0.3	22.5 ± 0.5	15.9 ± 0.1	15.0 ± 0.5	25.1 ± 0.1
	Lignin	21.1 ± 0.4	22.6 ± 1.7	21.4 ± 0.6	22.9 ± 0.3	24.4 ± 0.2	15.7 ± 0.4
5 g NaOH /100 g TS	Glucan	39.5 ± 0.4	40.2 ± 0.1	42.5 ±1.2	36.5 ± 0.2	47.2 ± 0.1	43.4 ± 0.3
	Xylan	23.3 ± 0.0	25.9 ± 0.2	24.4 ± 0.7	17.5 ± 1.8	16.1 ± 0.1	26.3 ± 0.3
	Lignin	18.8 ± 0.7	21.1 ± 0.3	19.8 ± 0.3	21.2 ± 0.2	22.8 ± 0.5	14.4 ± 0.1
10 g NaOH /100 g TS	Glucan	43.3 ± 0.5	41.8 ± 0.8	44.3 ±0.1	38.7 ± 0.2	48.1 ± 0.7	45.7 ±0.4
	Xylan	23.1 ±0.4	25.8 ± 0.4	24.7 ±0.4	17.6 ± 0.1	15.2 ± 0.2	27.2 ± 0.2
	Lignin	16.6 ± 1.2	17.9 ± 1.2	16.0 ± 0.6	19.4 ± 0.1	21.3 ± 0.4	11.2 ± 0.4
20 g Na0H /100 g TS	Glucan	48.7 ± 1.3	45.4 ± 0.1	49.4 ± 1.1	40.9 ± 0.5	49.4 ± 0.6	50.9 ± 0.8
	Xylan	23.5 ± 1.0	24.8 ± 0.2	24.3 ± 0.6	17.8 ± 0.9	15.3 ± 0.1	27.4 ± 0.1
	Lignin	14.1 ± 2.0	15.8 ± 0.5	13.2 ± 0.3	18.1 ± 0.7	20.4 ± 0.4	7.7 ± 0.0

percentages of potential glucose and xylose in the pretreated solids that were solubilized through hydrolysis. Glucose and xylose produced during pretreatment were not included.

Glucose conversion increased with NaOH loading for all biomasses with maximum occurring at the highest pretreatment level (20 g NaOH per 100 g TS). Differences between biomasses with respect to glucose conversion were maintained through all pretreatment levels, with minimum values for MHW and maximum for CS. This is in agreement with reports from the literature that alkaline pretreatments are significantly more effective for corn stover than for hardwood (Balan et al. 2009; Banergee et al. 2010). Xylose conversion also increased with NaOH loadings, and here too, differences between biomasses were maintained through levels of pretreatment. At each level of pretreatment, minimum xylose conversion was associated with

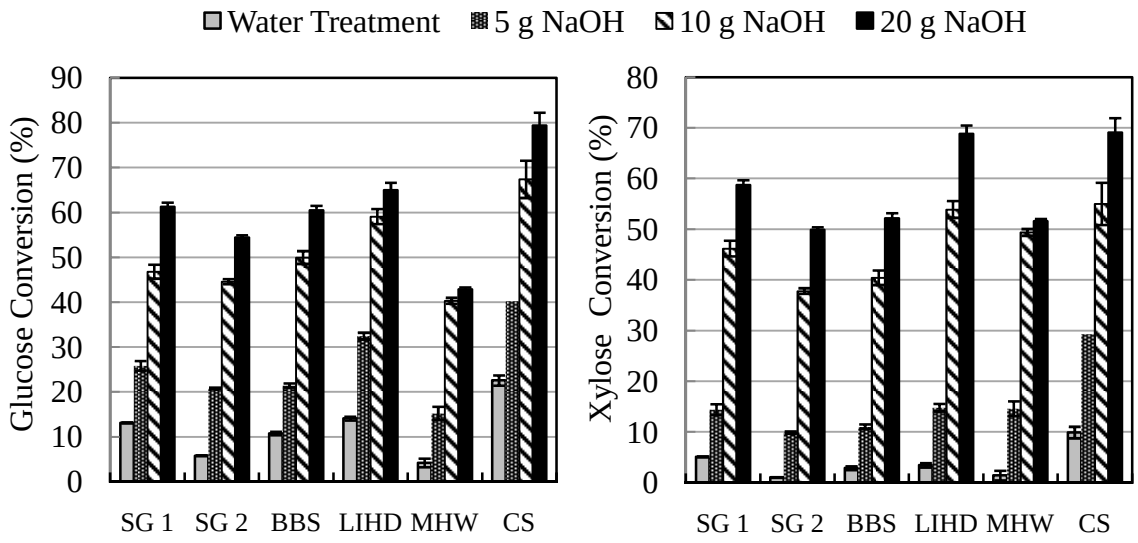

Figure 4.1 Glucose and xylose conversions from 72-h, cellulase- and β-glucosidase-mediated hydrolysis of SG1, SG2, BBS, LIHD, MHW, and CS pretreated with 0, 5, 10, and 20 g NaOH per 100 g TS (25 ºC, 24 h). Values are means of three replicates, and error bars represent standard deviations.

SG2 and maximum, with CS. Fig. 4.1 demonstrates that we generated 24 pretreated substrates with a wide range of glucose and xylose conversions from enzymatic hydrolysis, even though the glucan and xylan contents of the pretreated solids (Table 4.1) were relatively similar.

4.3.3 Development of PLS-Regression Models Based on FTIR Spectra

Sample FTIR-ATR spectra are shown in Fig. 4.2 for four samples of BBS pretreated with 0, 5, 10, and 20 g NaOH per 100 g TS. The similarity of the spectra illustrate the need for a chemometric technique (e.g. PLS regression) to extract information. PLS regression models were constructed from full FTIR spectra of the 24 pretreated biomass samples (X matrix) to predict glucose and xylose conversions (Y response vectors) [See section 4.2.6]. Models with 8 LVs resulted in the lowest cross-validation errors (RMSECV) of 7.8% for glucose and 9.3% for xylose and the highest Q^2 values of 0.86 for glucose and 0.84 for xylose. Fig. 4.3 presents plots of predicted versus measured values of glucose and xylose conversions for LOO cross-validation and model calibration of PLS models with 8 LVs. R^2 values for calibration were 0.99 and 0.98 for glucose and xylose conversions; and calibration errors (RMSEC) were 2.6 and 3.0%, respectively. R^2 and RMSEC describe how well the model fits the calibration data, while Q^2 and RMSECV represent how well the model will predict yields from new samples (Wold et al. 2001).

To assess the predictive ability of PLS models we considered values of Q^2 and RMSECV. Previous authors report that coefficient-of-determination (in our case, Q^2) values between 0.81 and 0.9 indicate good prediction, and values higher than 0.9 indicate excellent

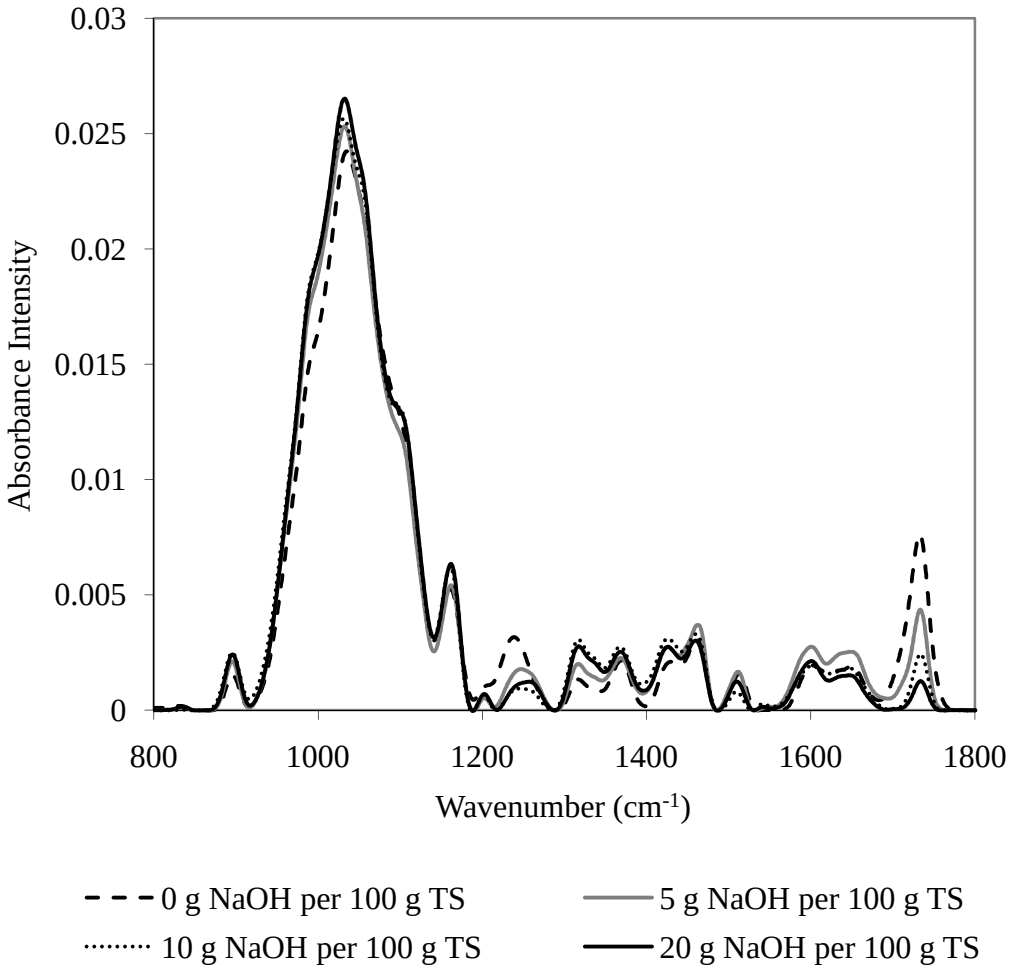

Figure 4.2 Averaged FTIR-ATR spectra (vector-normalized, ATR-corrected, and baseline-corrected) of big bluestem (BBS) grass pretreated with 0, 5, 10, and 20 g NaOH per 100 g TS.

prediction (Tamaki and Mazza 2011b). The difference between Q^2 and R^2 has also been considered for model assessment, with the suggestion that this difference should be less than 0.2 for models with good predictive ability (Chen et al. 2010). In the end, the usefulness of any model is determined by whether its prediction error is acceptable for whatever its intended use—a determination that is inevitably context-dependent.

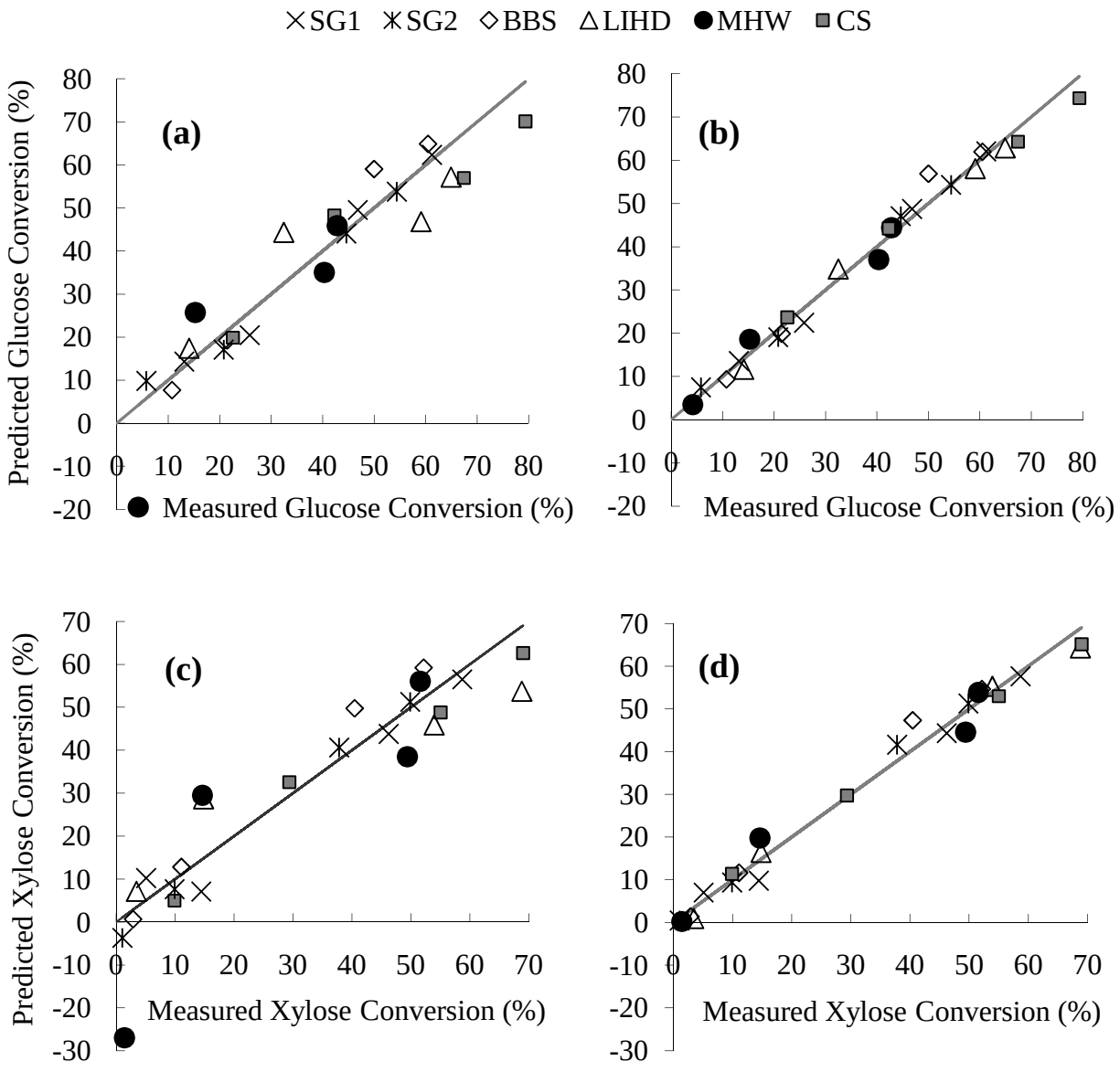

Figure 4.3 Predicted versus measured glucose (a and b) and xylose (c and d) conversions for validation (a and c) and calibration (b and d) of PLS regression models constructed from full spectra of SG1, SG2, BBS, LIHD, MHW, and CS pretreated with 0, 5, 10, and 20 g NaOH per 100 g TS. Glucose and xylose regression models were each constructed with 8 LVs.

Based on Q^2 values of 0.86 for glucose conversion and 0.84 for xylose conversion, and differences between R^2 and Q^2, which were 0.13 for glucose and 0.14 for xylose, we concluded that both models were of good predictive ability. However, even though these PLS models predict sugar conversions satisfactorily, validation plots [Figs. 4.3(a) and 4.3(c)] show that one sample— MHW pretreated with 0 g NaOH per 100 g TS — fit both regression models significantly worse than the remainder of the samples. This lack of fit along with PLS-score plots (plots not shown) suggest that removing this sample might improve PLS-regression models. However, the motivation for the present study is to predict sugar productions from a wide variety of biomasses. Hence, we chose to attempt improvement of PLS regression models — both in terms of their simplicity (i.e., low numbers of LVs retained in the model) and in terms of predictive ability — through variable selection, while retaining all 24 samples for modeling.

A variable was defined as unimportant if the 90% CI of its regression coefficient spanned the value of zero — i.e., its coefficient was neither significantly positive nor significantly negative. Variable selection reduced the number of variables from the 1039 wavenumbers of each full spectrum to 411 and 429 wavenumbers for modeling glucose and xylose conversions using truncated spectra. The regions of the IR spectrum retained after variable selection are shown in Fig. 4.4 for the final, truncated-spectra conversion models. Spectral interpretation is presented in a subsequent section.

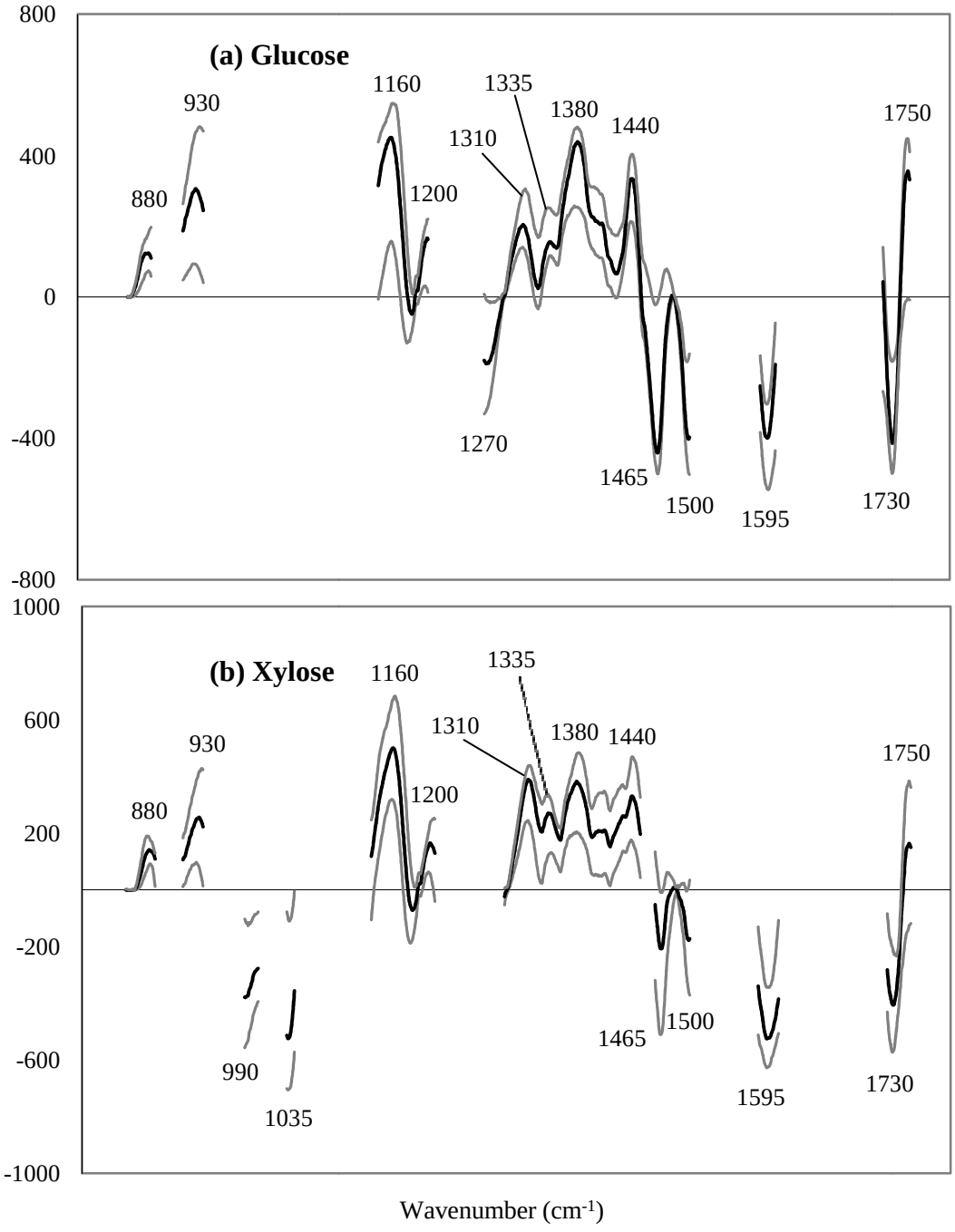

Figure 4.4 Regression-coefficient values versus wavenumber of the final PLS regression models for (a) glucose conversion and (b) xylose conversion. Both models were constructed from truncated spectra (after variable selection). Wavenumber values are listed above and below peaks in units of cm^{-1}. Units of the y–axis are arbitrary [sugar conversion/(absorbance intensity)].

4.3.4 Glucose and Xylose Conversion Models Using Truncated Spectra

Glucose- and xylose-conversion models with 4 LVs were chosen based on minimum RMSECV values. Validation plots, presented in Figs. 4.5 (a) and 4.5(c), show that variable selection improved model validation, as demonstrated by decreases in RMSECV values and increases in Q^2 values, along with simplifying the model, as demonstrated by the reduction in the number of LVs from 8 to 4. RMSECV values decreased from 7.8% to 6.8% for glucose and from 9.3% to 7.6% for xylose. Furthermore, values of Q^2 increased from 0.86 to 0.9 for glucose, and from 0.84 to 0.89 for xylose, indicating that the truncated models are of excellent (glucose) and good (xylose) predictive abilities.

Calibration plots presented in Figs. 4.5(b) and 4.5(d) [compared with those in Figs. 4.3(b) and 4.3(d)] show that variable selection resulted in increased RMSEC values and decreased R^2 values. This is to be expected. Calibration fits will inevitably worsen when fewer variables are used, but calibration plots should not be used alone to assess the predictive ability of multivariate-regression models. High RMSECV, relative to RMSEC, is characteristic of over-fitting, as is low Q^2 relative to R^2. In that regard, we note that variable selection resulted in decreases in the difference between R^2 and Q^2 for both glucose and xylose. This difference decreased from 0.13 to 0.04 for glucose, and from 0.14 to 0.05 for xylose, again demonstrating improved predictive ability of both regression models with variable selection. Additionally, cross-validation plots [Figs. 4.5(a) and 4.5(c)] show that variable selection eliminated the large validation error associated with MHW pretreated with 0 g NaOH per 100 g TS seen in the full-spectrum models [Figs. 4.3(a) and 4.3(c)]. Variable selection improved PLS regression models for glucose and xylose conversion without requiring the removal of sample-types.

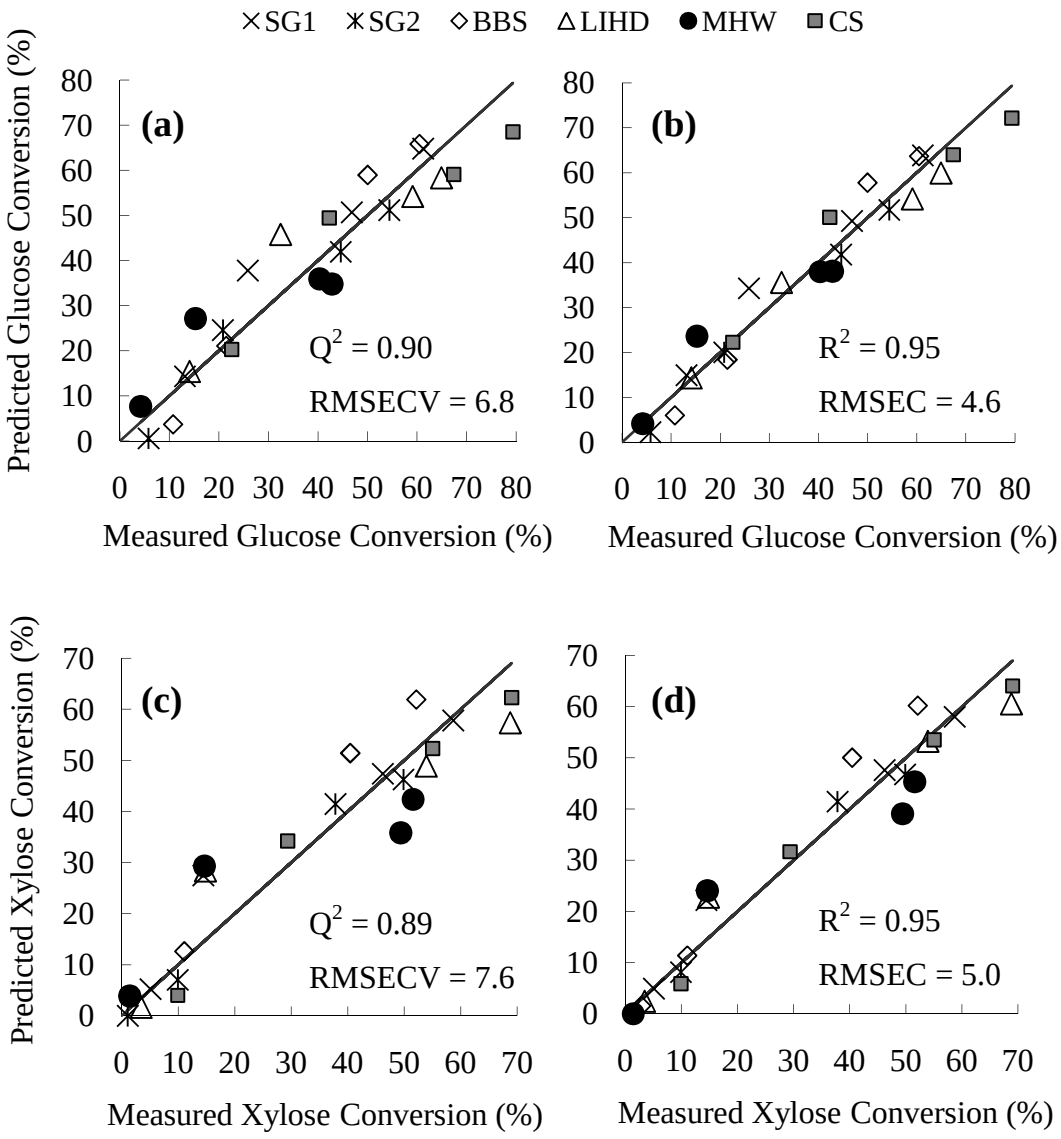

Figure 4.5 Predicted versus measured glucose (a and b) and xylose (c and d) conversions for validation (a and c) and calibration (b and d) of PLS regression models constructed from truncated spectra of SG1, SG2, BBS, LIHD, MHW, and CS pretreated with 0, 5, 10, and 20 g NaOH per 100 g TS. Glucose and xylose models were constructed with 4 LVs.

4.3.5 Glucose and Xylose Yield Models Using Truncated Spectra

We demonstrated that FTIR spectra of pretreated solids can satisfactorily predict sugar *conversions* from enzymatic hydrolysis. To predict sugar *yields* (g sugars/100 g TS) from sugar conversions requires, additionally, knowing the potential sugar contents per 100 g TS — values we measured on the pretreated solids using wet chemistry (Table 4.1). Since investigators have demonstrated that FTIR spectra can be used to estimate cellulose, hemicellulose, and lignin contents of pretreated biomasses (Tucker et al. 2001), this prompts a question: How well can FTIR spectra alone predict sugar yields from enzymatic hydrolysis of pretreated biomasses, without knowing their compositions — i.e., bypassing wet-chemistry analyses entirely?

Following the same procedures described above for sugar conversion models, we constructed PLS regression models for sugar yields. We began with full-spectrum models; examined the regression coefficient matrices; then proceeded with variable selection (i.e., spectra truncation). We again used glucose and xylose enzymatic-hydrolysis data from the 24 samples of six pretreated biomasses, this time in units of grams sugars per 100 g TS. The regions of the IR spectrum retained after variable selection (based on 90% CIs of the regression coefficients) are shown in Fig. 4.6. Spectral interpretation is presented in a subsequent section. Glucose- and xylose-yield models with 5 LVs were chosen based on minimum RMSECV values. It is reasonable that the yield models required one more LV than did the conversion models, since the FTIR spectra are effectively required to bear an additional burden in modeling yield: namely, the estimation of potential sugars in the solid substrates.

Cross-validation and calibration plots and statistical descriptors are presented in Fig. 4.7. RMSECV values were 3.2 for glucose and 1.9 for xylose, which are similar to the validation errors associated with the final sugar-conversion models, after accounting for the different

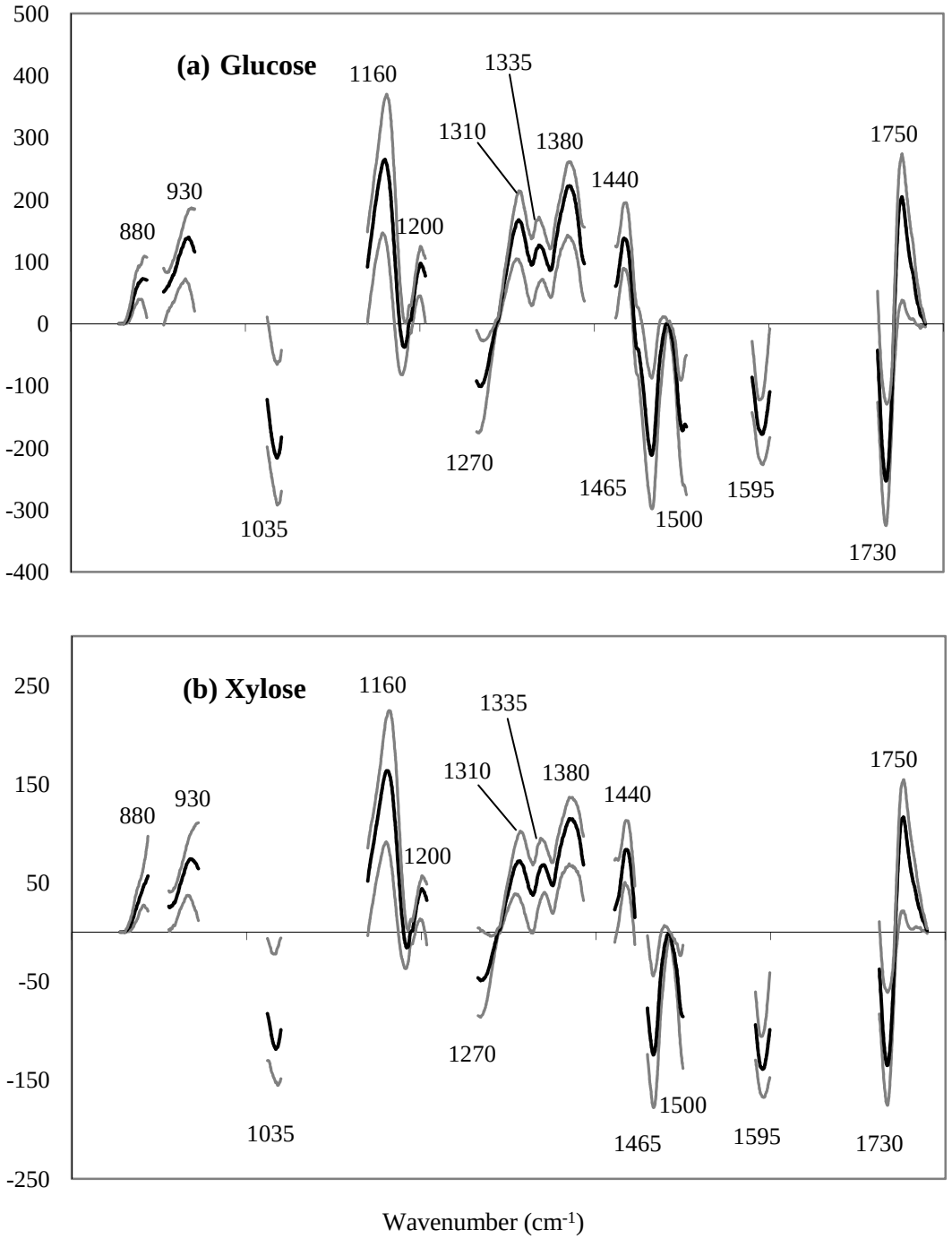

Figure 4.6 Regression-coefficient values versus wavenumber of the final PLS regression models for (a) glucose yield and (b) xylose yield. Both models were constructed from truncated spectra (after variable selection). Wavenumber values are listed above and below peaks in units of cm^{-1}. Units of the y–axis are arbitrary [sugar yield/(absorbance intensity)].

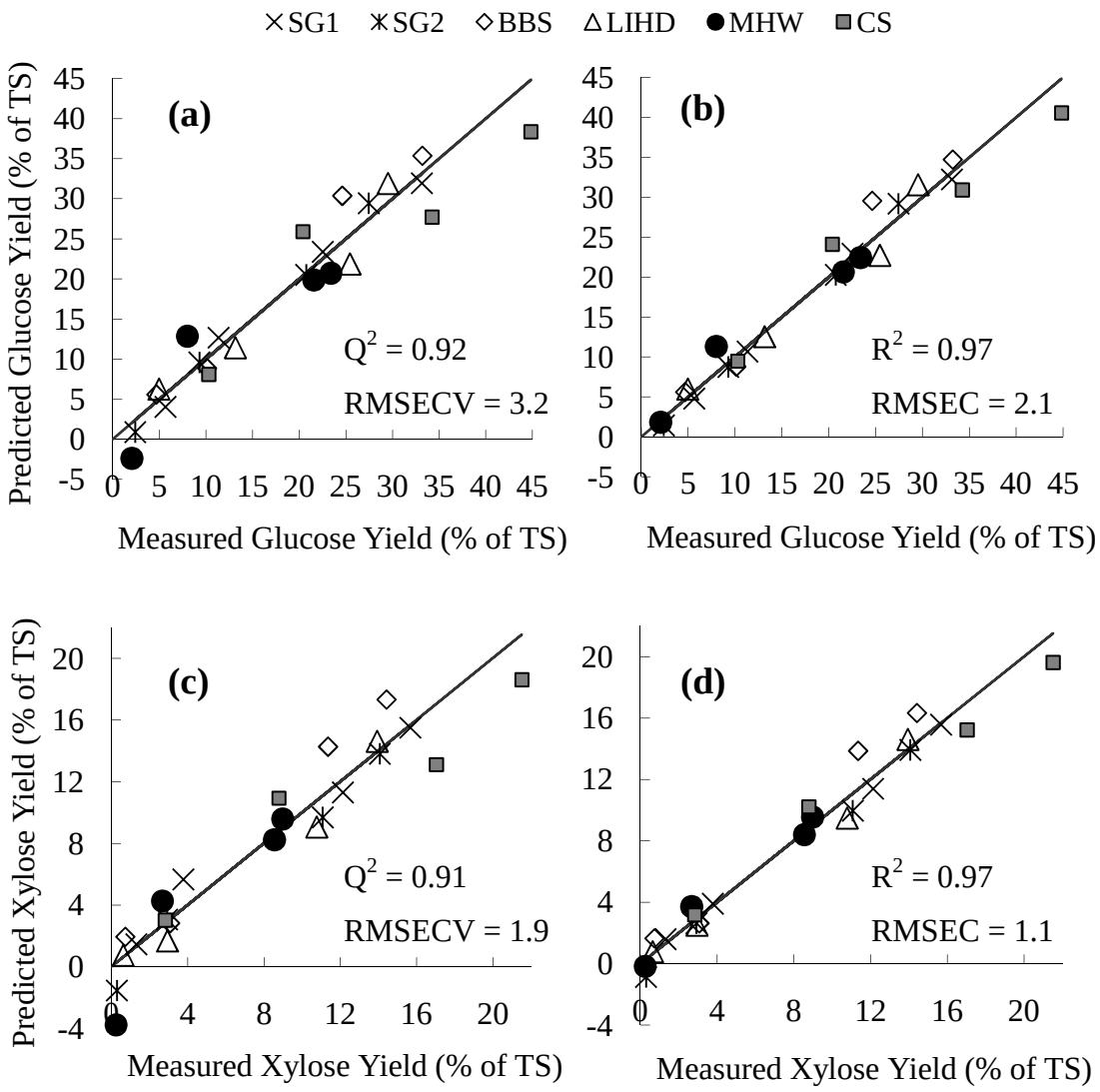

Figure 4.7 Predicted versus measured glucose (a and b) and xylose (c and d) yields for validation (a and c) and calibration (b and d) of PLS regression models constructed from truncated spectra of SG1, SG2, BBS, LIHD, MHW, and CS pretreated with 0, 5, 10, and 20 g NaOH per 100 g TS. Glucose and xylose models were constructed with 5 LVs.

units — i.e. sugar yields versus sugar conversions. Moreover, Q^2 values were 0.92 for glucose and 0.91 for xylose, and differences between Q^2 and R^2 were 0.05 for glucose and 0.06 for xylose, indicating that both sugar-yield models are of excellent predictive ability. The xylose validation plot [Fig. 4.7(c)], however, indicates that the model did not satisfactorily predict xylose yield from MHW pretreated with 0 g NaOH per 100 g TS, even after variable selection. This contrasts with the final xylose-conversion model, which predicted sugar conversion reasonably well for all samples [Fig. 4.5 (c)]. Still, the overall goodness-of-fit of glucose- and xylose-yield models suggests that FTIR can eliminate the need for wet-chemistry analyses and enzymatic assays used to estimate saccharification potential of pretreated biomass.

4.3.6 Analysis of Regression Coefficients

Regression coefficients with 90% CIs for glucose- and xylose-PLS models are presented in Figs. 4.4 and 4.6. Coefficients significantly greater than zero correlated positively with sugar conversions (Fig. 4.4) and sugar yields (Fig. 4.6); likewise, coefficients significantly less than zero correlated negatively with sugar conversions (Fig. 4.4) and sugar yields (Fig. 4.6). The four models constructed from truncated spectra were very similar to one another with respect to the presence and magnitude of spectral regions in final coefficient matrixes. Three differences were as follows: one peak at 990 cm^{-1} was present only in the xylose conversion model; a second peak at 1035 cm^{-1} was present in all models except for the glucose conversion model; and a third peak that spanned the region from1270 to 1280 cm^{-1} was present in all models except for the xylose-conversion model. Virtually all peaks included in the final coefficient matrices have been assigned previously to bonds and/or functional groups found in lignocellulosic polymers (Table

4.2), indicating that the models developed here are based on real chemical information and not chance correlation.

Three wavenumbers attributed to lignin — 1465, 1500, and 1595 cm^{-1} (Faix 1992; Owen and Thomas 1989) — correlated negatively with glucose and xylose conversions and yields, which is expected, since lignin is known to limit saccharification of pretreated lignocellulose. A peak of negative magnitude that spans 1270 and 1280 cm-^{1} was present in glucose- and xylose-yield models as well as the glucose conversion model. Absorbance at 1270 cm^{-1} is attributed to guaicyl lignin (Faix 1992) and absorbance at 1280 cm^{-1} is attributed to cellulose crystallinity (Hulleman et al. 1994). Since lignin and cellulose crystallinity limit enzymatic hydrolysis, the negative magnitudes of regression coefficients at 1270 to 1280 cm^{-1} make sense. However, we cannot explain why this region of the spectrum is not present in the xylose-conversion model.

Absorbance bands at 1730 cm^{-1} and 1750 cm^{-1} are associated with C=O bonds, and while coefficients in the region of 1730 cm^{-1} were negative, coefficients in the 1750 cm^{-1} range were positive. Owen and Thomas (1989) noted that absorbance peaks associated with C=O bonds vary from approximately 1720 to 1755 cm^{-1}, based on type of wood. Moreover, they attributed aldehydes and ketones to the lower end of this range, and free esters and acids to the upper end. The presence of aldehydes and ketones is indicative of intact hemicellulose and lignin, both known to limit cellulase-mediated hydrolysis (Grabber 2005; Ohgren et al. 2007); thus, negative magnitudes of coefficients at 1730 cm^{-1} make sense. The presence of free carboxyl groups, on the other hand, suggests ester-link breakage, previously correlated with increased digestibility of alkaline pretreated wheat straw (Pavlostathis and Gossett 1985); thus, positive coefficients at 1750 cm^{-1} are reasonable.

Table 4.2 Assignment of wavenumber bands present in the regression-coefficient matrices associated with final PLS models for glucose and xylose conversions and yields (Figs. 4.4 and 4.6).

Wavenumber cm^{-1}	Functional Group	Polymer
875	glycosidic linkage[1]	hemicellulose
930	glycosidic linkage[1]	cellulose, hemicellulose
990	cellulose I to II with alkali addtion[2]	cellulose
1035	C-O, and C=C. and C-C-O stretching[3,6]	cellulose, hemicellulose, lignin
1160	C-O-C asymmetrical stretching[4,5]	cellulose, hemicellulose
1200	O-H bending[5]	cellulose, hemicellulose
1270	aromatic ring vibration + C=O stretch[6]	guaicyl lignin[6]
1280	C-H bending[5]	crystalline cellulose[5]
1310	CH_2 wagging[2,5]	cellulose, hemicellulose
1335	C-H vibration[6], O-H in-plane bending[2]	cellulose, hemicellulose; lignin[6]
1380	C-H bending[4,6,7]	cellulose, hemicellulose, lignin
1440	O-H in-plane bending[4]	cellulose, hemicellulose, lignin
1465	C-H deformation[6]	lignin
1500	aromatic ring vibration[6]	lignin
1595	aromatic ring vibration + C=O stretch[6]	lignin
1730	ketone/aldehyde C=O stretch[6,8]	hemicellulose
1750	free esters and acids C=O stretch[8]	hemicellulose

[1]Kacuráková et al. 2000; [2]Nelson and O'Connor 1964a; [3]Wilson et al. 2000; [4]Marchessault 1962; [5]Hulleman et al. 1994; [6]Faix 1992 ; [7]Nelson and O'Connor 1964b; [8]Owen and Thomas 1989.

Wavenumbers between 800 and 1440 cm^{-1} are attributed to bonds present in multiple plant-cell-wall polymers [e.g. C-H bending in cellulose, hemicellulose, and lignin at 1380 cm^{-1} (Faix 1992; Nelson and O'Connor 1964a)]. Moreover, many absorption bands in this region result from vibrations of multiple bond types [e.g. 1335 cm^{-1} is attributed to C-H, C-O, and O-H bonds (Faix 1992; Hulleman et al. 1994; Nelson and O'Connor 1964b)], making this area of the spectrum extremely difficult to interpret (Faix 1992; Hulleman et al. 1994).

4.4 Conclusion

FTIR-ATR spectroscopy combined with PLS regression was used to estimate 72-h glucose and xylose conversions and yields from cellulase and β-glucosidase mediated hydrolysis of six types of lignocellulosic biomass that underwent four levels of alkaline pretreatment — a data set providing a broad range of conversions and yields. We were able to construct PLS-regression models that satisfactorily predicted glucose and xylose conversions (g sugars per 100 g potential sugars), as well as glucose and xylose yields (g sugars per 100 g TS). The ability to predict sugar yields from FTIR spectra of pretreated solids alone, without wet chemical analyses, suggests that this method could be very useful if the approach works equally well when applied to other biomasses and other pretreatment technologies.

Variable selection based on 90% CIs of individual variable coefficients improved the predictive ability of all four PLS models, while simplifying them (i.e. lowering the number of LVs required) and thus making them easier to interpret. Regions of the IR spectrum that were retained after variable selection could, in all instances, be identified with known chemical functional groups in lignocellulose. Spectral attributions, while tentative, indicate that the PLS models are based in chemical reality, and that FTIR spectra contain chemical information that

correlates with recalcitrance. Further development of this work may elucidate mechanisms by which chemical features of lignocellulose limit enzymatic hydrolysis of lignocellulosic biomass.

Acknowledgements

This study was supported with funding from the Department of Transportation and the Northeast Sun Grant Initiative. We thank Larry Walker and Stephane Corgié at the Biofuels Research Laboratory, Cornell University, for providing analytical equipment and instruction in FTIR analysis; Deborah Ross and Michael Van Amburgh, Cornell University, for their facilities and instruction in biomass characterization; the Nanobiotchnology Center at Cornell University for use of the FTIR; Bryan Hanson from DePauw University for instruction in R; Hillary Mayton from Cornell University, Matt McArdle from MESA Inc., James McMillan from NREL, Tom Stickle from Ligonier, PA, and Jason Hill and Matt Ziehr from the University of Minnesota for providing biomass samples; and Genencor Danisco for providing enzyme mixtures.

REFERENCES

Allison, GG, Thain, S.C., Morris, P. Morris, C., Hawkins, S., Hauck, B., Barraclough, T., Yates, N., Shield, I., Bridgwater, A., Donnison, I.S. 2009. Quantification of hydroxycinnamic acids and lignin in perennial forage and energy grasses by Fourier-transform infrared spectroscopy and partial least squares regression. Bioresour Technol 100:1252-1261.

Andersen CM, Bro R. 2010. Variable selection in regression — a tutorial. J Chemometrics 24:728-737.

Atalla RH, Brady JW, Matthews JF, Ding S-Y, Himmel ME. 2008. Structure of plant cell wall celluloses. In: Himmel ME, editor. Biomass recalcitrance: Deconstructing the plant cell wall for bioenergy. Oxford, United Kingdom: Blackwell Publisher Ltd. p 188-212.

Balan V, Sousa LC, Chundawat SPS, Marshall D, Sharma LN, Chambliss CK, Dale BE. 2009. Enzymatic digestibility and pretreatment degradation products of AFEX treated hardwoods (*Populus nigra*). Biotechnol Prog 25(2):365-375.

Banergee G, Car S, Scott-Craig JS, Borrusch MS, Walton JD. 2010. Rapid optimization of enzyme mixtures for deconstruction of diverse pretreatment/biomass feedstock combinations. Biotechnol Biofuels 3:1-15.

Casler MD, Jung H-JG. 2006. Relationships of fibre, lignin, and phenolics to in vitro fibre digestibility in three perennial grasses. Anim Feed Sci Technol 125:151-161.

Chandra RP, Bura R, Mabee WE, Berlin A, Pan X, Saddler JN. 2007. Substrate pretreatment: The key to effective enzymatic hydrolysis of lignocellulosics? Adv Biochem Eng Biotechnol 108:67-93.

Chen H, Ferrari C, Angiuli M, Yao J, Raspi C, Bramanti E. 2010. Qualitative and quantitative analysis of wood samples by Fourier transform infrared spectroscopy and multivariate analysis. Carbohydr Polym 82:772-778.

Chong IG, Jun CH. 2005. Performance of some variable selection methods when multicollinearity is present. Chemometrics Intelligent Lab Syst 78:103-112.

Corgié SC, Smith HM, Walker LP. 2011. Enzymatic transformations of cellulose assessed by quantitative high-throughput Fourier transform infrared spectroscopy (QHT-FTIR). Biotechnol Bioeng 108:1509-1520.

de Jong S. 1993. SIMPLS: an alternative approach to partial least squares regression. Chemometrics Intelligent Lab Syst 18:251-263.

Efron B, Tibshirani R. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1:54-75.

Faix O. 1992. Fourier Transform Infrared Spectroscopy. In: Lin SY, Dence CW, editors. Methods in lignin chemistry. Berlin: Springer-Verlag. p 83-109.

Farrell AE, Plevin RJ, Turner BT, Jones AD, O'Hare M, Kammen DM. 2006. Ethanol can contribute to energy and environmental goals. Science 311.

Geladi P, Kowalski BR. 1986a. An example of 2-block predictive partial least-squares regression with simulated data. Anal Chim Acta 185:19-32.

Geladi P, Kowalski BR. 1986b. Partial least-squares regression: a tutorial. Anal Chim Acta 185:1-17.

Gilbert HJ, Stålbrand H, Brumer H. 2008. How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. Curr Opin Plant Biol 11:1-11.

Gollapalli LE, Dale BE, Rivers DM. 2002. Predicting digestibility of ammonia fiber explosion (AFEX)-treated rice straw. Appl Biochem Biotechnol 98:23-35.

Grabber JH. 2005. How do lignin composition, structure, and cross-linking affect degradability? A review of cell wall model studies. Crop Sci 45:820-831.

Gwon JG, Lee SY, Doh GH, Kim JH. 2010. Characterization of chemically modified wood fibers using FTIR spectroscopy for biocomposites. J Appl Polym Sci 116:3212-3219.

Hulleman SHD, van Hazendonk JM, van Dam JEG. 1994. Determination of crystallinity in native cellulose from higher plants with diffuse reflectance Fourier infrared spectroscopy. Carbohydr Res 261:163-172.

Jørgensen H, Kristensen JB, Felby C. 2007. Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. Biofuel Bioprod Bior 1:119-134.

Kacuráková C, Capek P, Sasinkova V, Wellner N, Ebringerova A. 2000. FT-IR study of plant cell wall model compounds: pectic polysaccharides and hemicelluloses. Carbohydr Polym 43:195-203.

Kristensen JB, Thygesen LG, Felby C, Jørgensen H, Elder T. 2008. Cell-wall structural changes in wheat straw pretreated for bioethanol production. Biotechnol Biofuels 1. Available at http://www.biotechnologyforbiofuels.com/1754-6834/1/5

Marchessault, RH. 1962. Application of infra-red spectroscopy to cellulose and wood polysaccharides. Pure Appl Chem 5:107-129.

Merino ST, Cherry J. 2007. Progress and challenges in enzyme development for biomass utilization. Adv Biochem Eng Biotechnol 108:95-120.

Mevik BH, Wehrens R. 2007. The pls package: principal component and partial least squares regression in R. J Stat Softw 18:1-24.

Nelson M, O'Connor RT. 1964a. Relation of certain infrared bands to cellulose crystallinity and crystal lattice type. Part I. Spectra of lattice types I, II, III, and of amorphous cellulose. J Appl Polym Sci 8:1311-1324.

Nelson M, O'Connor RT. 1964b. Relation of certain infrared bands to cellulose crystallinity and crystal lattice type. Part II. A new infrared ratio for estimation of crystaillinity in celluloses I and II. J Appl Polym Sci 8:1325-1341.

Ohgren K, Bura R, Saddler J, Zacchi G. 2007. Effect of hemicellulose and lignin removal on enzymatic hydrolysis of steam pretreated corn stover. Bioresour Technol 98:2503-2510.

Owen NL, Thomas DW. 1989. Infrared studies of "hard" and "soft" woods. Appl Spectrosc 43:451-455.

Pandey KK, Pitman AJ. 2003. FTIR studies of the changes in wood chemistry following decay by brown-rot and white-rot fungi. Int Biodeterior Biodegrad 52:151-160.

Pavlostathis SG, Gossett JM. 1985. Alkaline treatment of wheat straw for increasing anaerobic biodegradability. Biotechnol Bioeng 27:334-344.

Raiskila S, Pulkkinen M, Laakso T, Fagerstedt K, Löija M, Mahlberg R, Paajanen L, Ritschkoff A-C, Saranpää P. 2007. FTIR spectroscopic prediction of Klason and acid soluble lignin variation in Norway spruce cutting clones. Silva Fennica 41:351-371.

Saha BC. 2003. Hemicellulose bioconversion. J Ind Microbiol Biotechnol 30:279-291.

Sills DL, Gossett JM. 2011. Assessment of commercial hemicellulases for saccharification of alkaline pretreated perennial biomass. Bioresour Technol 102:1389-1398.

Sluiter A, Hames B, Ruiz R, Scarlata C, Sluiter J, Templeton D, Crocker D. 2008. Determination of structural carbohydrates and lignin in biomass. Golden, CO: National Renewable Energy Laboratory. NREL Laboratory Procedure.

Somerville C. 2006. The billion-ton biofuels vision. Science 312:1277.

Tamaki Y, Mazza G. 2011. Rapid determination of lignin content of straw using attenuated total reflectance Fourier transform mid-infrared spectroscopy. J Agric Food Chem 59:504-512.

Tilman D, Hill J, Lehman C. 2006. Carbon negative biofuels from low-input high-diversity grassland biomass. Science 314:1598-1600.

Tucker MP, Nguyen QA, Eddy FP, Kadam KL, Gedvilas LM, Webb JD. 2001. Fourier transform infrared quantitative analysis of sugars and lignin in pretreated softwood solid residues. Appl Biochem Biotechnol 91-93:51-61.

Van Soest P, Robertson J, Lewis B. 1991. Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. J Dairy Sci 74:3583-3597.

Wehrens R. 2011. Chemometrics with R. Multivariate data analyis in the natural sciences and life sciences. Gentleman R, Hornik K, Parmigiani G, editors. Berlin: Springer-Verlag

Wilson R, Smith AC, Kacuráková M, Wellner N, Waldron W. 2000. The mechanical properties and molecular dynamics of plant cell wall polysaccharides studied by Fourier-transform infrared spectroscopy. Plant Phys 124:397-405.

Wold S, Sjostrom M, Eriksson L. 2001. PLS-regression: a basic tool of chemometrics. Chemometrics Intelligent Lab Syst 58:109-130.

Wyman CE. 2007. What is (and is not) vital to advancing cellulosic ethanol. Trends Biotechnol 25:153-157.

Zhou G, Taylor G, Polle A. 2011. FTIR-ATR-based prediction and modeling of lignin and energy contents reveals independent intra-specific variation of these traits in bioenergy poplars. Plant Methods 7(9). Available at http://www.plantmethods.com/content/7/1/9

CHAPTER 5

USING FTIR SPECTROSCOPY TO PREDICT EFFECTS OF ALKALINE PRETREATMENT ON SIX LIGNOCELLULOSIC BIOMASSES

Abstract

Fourier transform infrared, attenuated total reflectance (FTIR-ATR) spectroscopy, combined with partial least squares (PLS) regression, accurately predicted solubilization of plant cell wall constituents and NaOH consumption through pretreatment and overall sugar productions from combined pretreatment and enzymatic hydrolysis. PLS regression models were constructed by correlating FTIR spectra of six raw biomasses — two switchgrass cultivars, big bluestem grass, a low-impact, high-diversity mixture of prairie biomasses, mixed hardwood, and corn stover — plus alkali loading in pretreatment to nine dependent variables: glucose, xylose, lignin, and total solids solubilized in pretreatment; NaOH consumed in pretreatment; and overall glucose and xylose conversions and yields from combined pretreatment and enzymatic hydrolysis. PLS models predicted the dependent variables with the following values of coefficient of determination for cross-validation (Q^2): 0.86 for glucose, 0.90 for xylose, 0.79 for lignin, and 0.85 for TS solubilized in pretreatment; 0.83 for alkali consumption; 0.93 for glucose conversion, 0.94 for xylose conversion, and 0.88 for glucose and xylose yields. The sugar yield models are noteworthy for their ability to predict overall saccharification through combined pretreatment and enzymatic hydrolysis per mass dry untreated solids without *a priori* knowledge of the composition of solids. All peaks with significant variable-important-for-projection (VIP) scores have been attributed to chemical features of lignocellulose, demonstrating the models were based on real chemical information. These models suggest that PLS regression can be

applied to FTIR-ATR spectra of raw biomasses to rapidly predict effects of pretreatment on solids and on subsequent enzymatic hydrolysis.

Keywords: lignocellulose, pretreatment, FTIR, PLS regression

5.1 Introduction

Lignocellulose, the material that makes up plant cell walls, is the largest source of renewable carbon on earth (Pauly and Keegstra 2008). Current concerns over climate change have renewed interest in transforming plant cell walls into simple sugars that can be further processed into valuable fuels and chemicals (Dodds and Gross 2007; Somerville 2006). The most promising saccharification technology consists of low-severity chemical pretreatment followed by cellulase- and hemicellulase-mediated hydrolysis (Faust et al. 2008). However, plants cell walls have evolved to resist biological depolymerization, and the recalcitrance of their cell walls limits sugar production from combined pretreatment and enzymatic hydrolysis (Himmel et al. 2007).

Lignocellulose is composed of three main constituents: cellulose and hemicellulose (two types of carbohydrate polymers) and lignin (a phenolic polymer). Recalcitrance of lignocellulose is a result of chemical interactions among carbohydrate polymers and interactions between carbohydrate polymers and lignin, which together create a tightly bound, impenetrable material. The hemicellulose-lignin matrix sheaths carbohydrate bonds, limiting accessibility of cell-wall degrading enzymes, including cellulases, hemicellulases, and esterases. Additionally, the chemical make-up of cell walls varies widely across plant species, plant tissues, and even within single plant cells (Pauly and Keegstra 2010), making the design and optimization of pretreatment

and enzymatic hydrolysis processes challenging (Chandra et al. 2007; Merino and Cherry 2007). Differences in the type of cellulose (e.g. I or II), hemicellulose (e.g. glucuronoarabinoxylan or xyloglucan), and lignin (syringyl or guaicyl), as well as the bonds that link them with one another, influence recalcitrance. The mechanisms by which these differences limit saccharification of lignocellulose are poorly understood.

Wet-chemistry analyses used to measure the composition of lignocellulose, and, thus, to determine changes in composition through pretreatment, do not elucidate structural and chemical details that influence extents of conversion of plant cell walls into monosaccharides through subsequent enzymatic hydrolysis. In the previous chapter, we demonstrated that Fourier transform infrared – attenuated total reflectance (FTIR–ATR) spectra of pretreated biomasses, combined with partial least squares (PLS) regression, satisfactorily predicted sugar productions from enzymatic hydrolysis of alkaline pretreated biomasses. This prompted the following question: Can FTIR spectra of raw biomasses, plus alkali loading in pretreatment, predict solubilization of biomass components, alkali consumption, and overall sugar production through pretreatment and enzymatic hydrolysis?

A number of researchers have constructed PLS regression models that predict chemical composition of lignocellulosic substrates from FTIR spectra (Allison et al. 2009a; Allison et al. 2009b; Raiskila et al. 2007; Tamaki and Mazza 2011a; Tamaki and Mazza 2011b; Zhou et al. 2011). However, even though it is possible to use different types of predictor variables in a single PLS model (Eriksson et al. 2001), there are no reports in the literature of chemometric models that use FTIR spectra of lignocellulosic materials, combined with other types of data (e.g. chemical dose used in processing), as predictor variables.

In the present study, six biomass substrates — two switchgrass cultivars; big bluestem grass; a low-impact, high-diversity mixture of prairie biomasses (Tilman et al. 2006), mixed hardwood; and corn stover — that represent a variety of potential plant-biomass feedstocks were subjected to four levels of alkaline pretreatment (0, 5, 10, and 20 g NaOH per 100 g TS), followed by enzymatic hydrolysis. FTIR spectra of raw biomasses and levels of pretreatment were used as predictor variables in two groups of PLS regression models. The first group of models predicted the effect of alkaline pretreatment on biomass components and NaOH consumption during pretreatment. The second group of models predicted overall glucose and xylose productions through combined pretreatment and enzymatic hydrolysis.

5.2 Materials and Methods

5.2.1 *Biomass Samples*

The six biomasses used in this study were described in the previous chapter (Section 4.2.1). Briefly, samples consisted of two switchgrass cultivars (SG1 and SG2); big bluestem grass (BBS); a low-impact, high-diversity mixture of 32 species of prairie biomasses (LIHD); mixed hardwood [MHW]; and corn stover [CS]. All biomasses were milled to pass a ¼-inch mesh screen.

5.2.3 Chemicals and Enzymes

Glucose and xylose were from Sigma-Aldrich, 98+% purity. Spezyme CP® cellulase (60 FPU/mL, 118 mg protein/mL) was provided by Genencor Danisco® US Inc. (Rochester, NY, USA). Novozyme® 188 ß-glucosidase (280 CBU/mL, 140 mg protein/mL) was from Sigma Aldrich (St. Louis, MO, USA).

5.2.4 Compositional Analysis

Glucan, xylan, and Klason lignin in solid samples (both before and after pretreatment) were quantified using a modification of the standard NREL LAP protocol (http://www.nrel.gov/biomass/analytical_procedures.html; Sills and Gossett 2011). Resulting glucose and xylose in acid hydrolysates were quantified via high performance liquid chromatography (HPLC) (Sills and Gossett, 2011). These were considered to be the potential glucose and xylose in the pretreated solids and were used to calculate glucose and xylose solubilized through pretreatment, as described in the following section. For purposes of describing solids compositions, they were converted to equivalent, polymeric glucan and xylan using conversion factors of 0.9 and 0.88, respectively. Glucose, xylose, and lignin contents of untreated biomasses are reported in Table 5.1.

Table 5.1 Composition of raw biomasses, % of dry matter (TS).

Biomass Type	Glucan[a]	Xylan[a]	Lignin[a]
Switchgrass 1	36.2 ± 0.2	20.1 ± 0.4	18.9 ± 1.0
Switchgrass 2	35.0 ± 0.3	23.5 ± 0.1	21.5 ± 0.1
Big Bluestem	37.9 ± 0.4	21.1 ± 0.4	19.9 ± 0.3
LIHD	30.9 ± 0.2	15.8 ± 0.0	23.9 ± 0.2
Mixed Hardwood	42.8 ± 0.2	14.6 ± 0.2	23.9 ± 0.2
Corn Stover	39.2 ± 0.7	23.2 ± 0.7	13.5 ± 0.7

[a]Mean values of three replicates ± standard deviations.

5.2.5 Pretreatment and Enzymatic Hydrolysis

Pretreatments of all biomasses were conducted at the following alkali levels: 0, 5, 10, and 20 g NaOH per 100 g TS with 5% (w/w) total solids concentration at 25 °C in batch reactors on a rotary shaker at 200 rpm for 24 hours as described previously (Sills and Gossett 2011). Solids were separated by filtering neutralized slurries through Miracloth (Calbiochem, San Diego, CA) prior to enzymatic hydrolysis. Additional triplicate pretreatment samples were conducted for the purpose of measuring TS solubilized in pretreatment. Pretreated samples were weighed after drying for 12 h at 105 ºC, followed by 2 hours in a desiccator. Values of glucose, xylose, and lignin solubilized in pretreatment were calculated using the results of solids compositional analysis combined with values of TS lost in pretreatment.

Enzymatic hydrolysis was conducted on pretreated solids with a modification of the standard NREL LAP protocol (http://www.nrel.gov/biomass/analytical_procedures.html; Sills and Gossett 2011). Briefly, hydrolysis was performed at 50 ºC in batch reactors on a rotary shaker at 130 rpm with a solids concentration of 2.5% TS (w/w) in 0.05 M citrate buffer (pH 4.8) with tetracycline (30 mg/L) and cyclohexamide (20 mg/L) to prevent microbial growth. Triplicate reactions were run with cellulase (Spezyme CP) and β-glucosidase (Novozyme 188) at loadings of 15 and 6 mg protein per g TS, respectively, corresponding to approximately 15 FPU + 25 CBU per g glucan. Triplicate substrate blanks without enzyme addition and enzyme blanks without substrate addition were run in parallel. Glucose and xylose were measured with HPLC (Sills and Gossett, 2011).

5.2.6 Calculation of Overall Sugar Conversion and Yields

We calculated two expressions of overall saccharification for each sugar: (1) glucose (or xylose) *conversion*, defined as the percentage of the potential glucose (or xylose) present in the raw biomass solubilized in pretreatment plus enzymatic hydrolysis; and (2) glucose (or xylose) *yield*, defined as the g glucose (or xylose) solubilized through pretreatment plus enzymatic hydrolysis per 100 g of raw TS. The former quantifies the accessibility of the sugar components in the raw biomass; the latter is perhaps a more useful parameter for the practitioner.

5.2.7 FTIR-ATR Spectroscopy

Prior to FTIR analysis, samples were dried under vacuum at 50 ºC for 24 h. Triplicate spectra of pretreated solids samples were collected by attenuated total reflectance (ATR) using a Vertex 80 FTIR spectrometer (Bruker Optics, Germany) equipped with a MIRacle-ATR sampling device (Pike Technologies, Madison, WI). Samples were pressed against a single-reflection diamond crystal with a torque knob that ensured identical pressure was applied to all samples. Spectra were obtained by averaging 64 scans from 4000–400 cm^{-1} at a resolution of 2 cm^{-1}, and corrected for background absorbance by subtraction of the spectrum of the empty ATR crystal. The ATR sample chamber was purged with N_2 and held under vacuum during sampling, and background samples were taken every 15 minutes. Spectra were vector-normalized, ATR-corrected, and baseline-corrected using a concave rubber-band correction to account for penetration depth and frequency variations using the OPUS software (Bruker, Germany). Triplicate spectra were averaged and converted into text files in OPUS for further analysis.

5.2.8 Partial Least Squares Regression Modeling

All data manipulation and statistical analyses were conducted with the software package R 2.9.2 (R Core Development Team, Vienna, Austria, 2009). Averaged FTIR spectra in the "fingerprint" region of 800-1800 cm^{-1} were combined into a single matrix with dimensions of 6 samples by 1039 wavenumbers as described in the previous chapter (Section 4.2.6). We then added an additional variable, "pretreatment level" to each sample vector, to construct predictor-variable matrices (X). X matrices were used in PLS regression models to predict response vectors (Y). For example, in a pretreatment model constructed to predict TS solubilization, the Y vector consisted of the %TS solubilizations from the 24 samples (6 biomasses x 4 pretreatment levels).

5.2.8.1 Variable Scaling and Transformation

One complication is that the predictor variable for pretreatment (NaOH level) has a significantly different magnitude than the FTIR predictor variables (wavenumber intensities). Often, when predictor variables in PLS regression models are of greatly different magnitudes, variables are auto-scaled, making all variables have a mean of zero and a variance of one (Geladi and Kowalski 1986). However, autoscaling is not recommended for spectral data because it causes spectral noise and actual peaks to be of equal magnitudes (Eriksson et al. 2001; Wehrens 2011). Moreover, Eriksson et al. (2001) reported that when two types of variables (e.g. spectral data and NaOH loading) have vastly different numbers of variables (e.g. 1039 spectral variables and one variable associated with NaOH loading) "block scaling" should be conducted (i.e. each variable type should be scaled differently), such that the model accurately represents the importance of both variable types. Ultimately, when predictor variables are of different

magnitudes, it is up to the modeler to determine appropriate scaling, and more often than not, proper scaling is determined via trial and error (Eriksson et al. 2001; Wold et al. 2001).

We investigated multiple scaling regimes to make alkali-loading data compatible with the spectral data. Fortunately, scaling of alkali loadings did not significantly change the fit of the regression models; however, scaling did change the relative magnitudes of the regression coefficients. In the end, we settled on the following method by which to scale the variable expressing alkali level, in constructing models to predict solubilization of TS, glucose, xylose, and lignin from pretreatment: Prior to building the X matrix, the four levels of NaOH loading were scaled such that the magnitude of the highest loading (20 g NaOH per 100 g TS) was 0.02, which was very similar to the maximum wavenumber intensity observed (0.021). In other words pretreatment levels were incorporated into the X matrix with the following values: 0, 0.005, 0.01, and 0.02.

Mere *scaling* of NaOH level was insufficient in PLS modeling of alkali consumption from pretreatment, and in modeling of overall (i.e., pretreatment + enzymatic hydrolysis) glucose and xylose conversions and yields from FTIR spectra of the raw biomasses and pretreatment levels. Because the effects of NaOH loading on these response variables were nonlinear, we found it necessary to perform a mathematical *transformation* of the NaOH-loading variable in these instances. After exploring a number of transformations (e.g. fourth root and natural logarithm) that did not improve model fits, we transformed values for alkali loading using the following logit-type expression, (Keer and Meador 1995)

$$Transformed\ PT\ Level = \ln\left(\frac{PT\ Level}{PT\ Level - 1}\right) \qquad (5.1)$$

where pretreatment levels were entered as g NaOH per 100 g TS.

The logit-type mathematical transformation did not allow us to incorporate data of 0 g NaOH per 100 g TS. Thus, we were not able to devise a transformation that both accommodated a pretreatment level of zero and resulted in a model that fit sugar conversions and yields reasonably well. In any event, the zero-level of pretreatment is probably not a level of practical interest.

5.2.8.2 Model Construction and Implementation

After variable scaling or variable transformation, two predictor variable matrices (X) were constructed. One X matrix was constructed for modeling the following dependent variables (Y vectors): glucose, xylose, lignin, and TS solubilized in pretreatment. For these models, four levels of pretreatment (scaled as described above) were added to each of the six sample vectors (each a spectrum of a specific biomass), creating four predictor vectors for each biomass, each with 1040 variables. The resulting X matrix had dimensions of 24 samples by 1040 predictor variables (1039 wavenumber intensities + pretreatment level). A second X matrix was constructed similarly for modeling alkali consumption as well as glucose and xylose conversions and yields (Y vectors). In this case, three levels of pretreatment (transformed as described in Eq. (5.1)) were added to each sample vector (or raw biomass spectrum), resulting in an X matrix with dimensions of 18 samples by 1040 variables (1039 wavenumbers + pretreatment level).

PLS regression models were developed as described in the previous chapter (Section 4.2.6) using the *pls* package (Mevik and Wehrens 2007). Briefly, mean-centered X matrices were correlated to mean-centered Y vectors using the SIMPLS (de Jong 1993) or OSCORES

(Martens and Naes 1989) algorithms. Note that the two algorithms resulted in identical models. Cross-validation was conducted using the leave-one-out (LOO) procedure to obtain a root mean square error of cross validation (RMSECV) and a coefficient of determination for validation (Q^2) (Geladi and Kowalski 1986; Stone 1974; Wehrens 2011), as described in the previous chapter (Section 4.2.7). The number of latent variables (LVs) retained in each model was chosen based on minimum values of RMSECV, which in all cases, corresponded to maximum values of $Q^{2.}$ The root mean square error of calibration (RMSEC) and the coefficient of determination for calibration (R^2) were then calculated for each model.

Ninety percent confidence intervals (CIs) of regression coefficients were calculated using a bootstrapping protocol, as described in previous chapter [Section 4.2.6] (Efron and Tibshirani 1986; Wehrens 2011). We considered variables with 90% CIs that did not span the value of zero to be important for predicting Y. To aid in spectral interpretation we also calculated the variables important for projection (VIP) scores (Chong and Jun 2005). VIP scores describe the importance of variables for explaining differences between samples with respect to the X matrix as well as for predicting the dependent variable (Y). A more detailed discussion of VIP scores is found in Section 2.4.2.3.

5.3 Results

5.3.1 Effect of NaOH Loading on Solubilization of Biomass Components

The effect of pretreatment on solubilization of biomass components — the percent of glucose (a), xylose (b), lignin (c), and TS (d) solubilized through pretreatment — is presented in Fig. 5.1. We were interested in how pretreatment affects component solubilization for two

reasons. First, it has been shown that solubilization of cell wall material can improve accessibility of cell wall degrading enzymes to the remaining carbohydrate substrates.

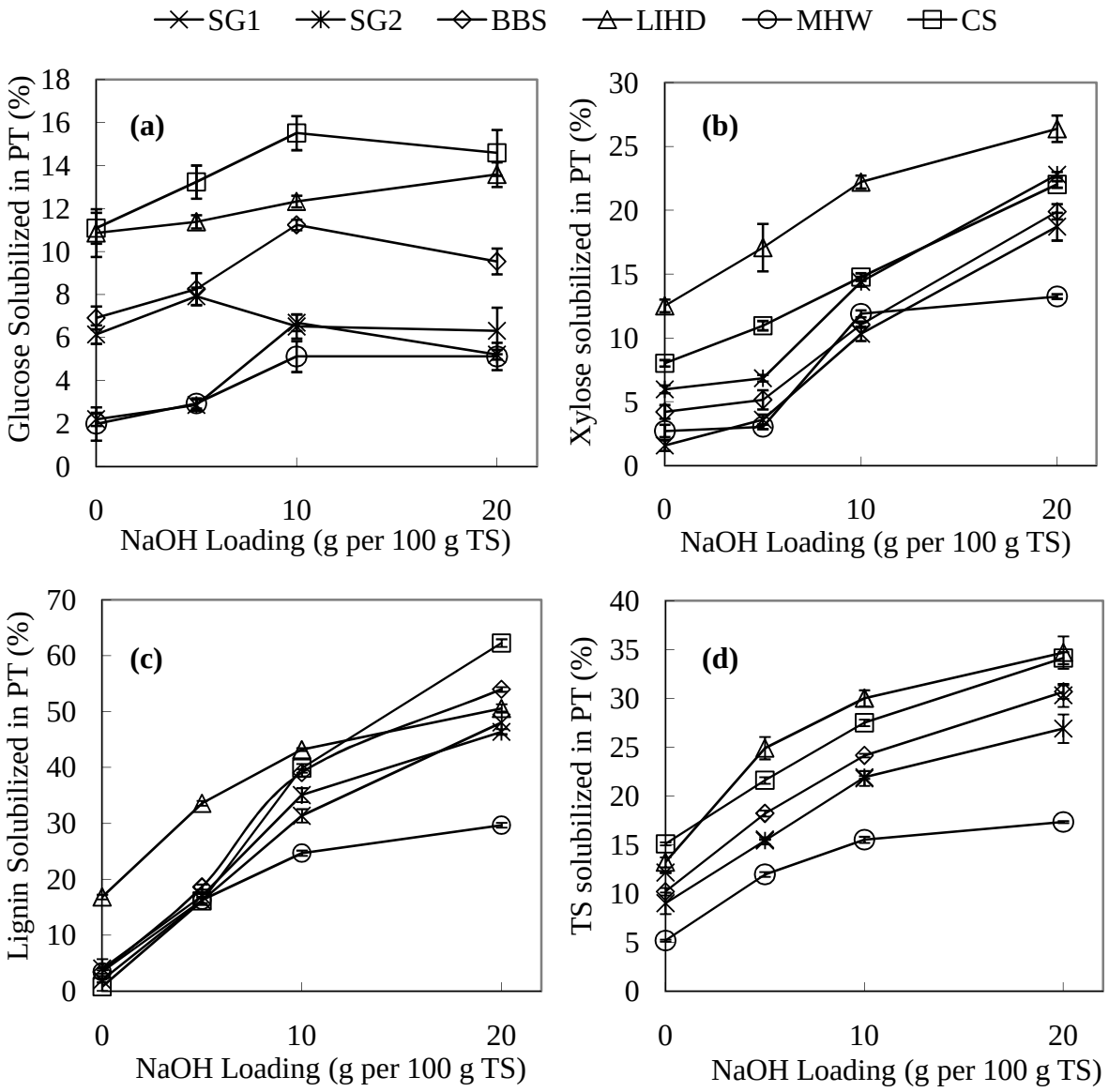

Figure 5.1 Effect of NaOH loading on glucose (a), xylose (b), lignin (c), and TS (d) solubilized in pretreatment for SG1, SG2, BBS, CS, LIHD, MHW, and CS. Values are averages of triplicate measurements and error bars represent standard deviations.

Solubilizations of hemicellulose and lignin (Ohgren et al. 2007; Mooney et al. 1998; Yang and Wyman 2004), in particular, have been correlated with increased extents of saccharification and digestibilities of pretreated plant substrates. Additionally, from a practical standpoint, we accounted for soluble sugars produced in pretreatment and added them to sugars produced through enzymatic hydrolysis to calculate overall conversions and yields. Ultimately, these data were collected to construct PLS regression models that predict the effect of pretreatment on component solubilizations and overall sugar conversions and yields from FTIR spectra of raw biomasses. The models are presented in a Section 5.3.6.

For all biomasses, NaOH loading had only a small effect on glucose solubilization [Fig. 5.1(a)]. At the highest level of pretreatment (20 g NaOH per 100 g TS), values of glucose solubilization ranged from 5% (SG2 and MHW) to 15% (CS). This agrees with reports in the literature that after low-temperature NaOH pretreatment, the majority of the original glucan is retained in pretreated solids (Pavlostathis and Gossett 1985; Zhao et al. 2008).

Solubilization of xylose increased with NaOH loading for all biomasses with maximum values observed at the highest level of pretreatment [Fig. 5.1(b)]. At the highest alkali loading of 20 g NaOH per 100 g TS, xylose solubilization ranged from 13% (MHW) to 26% (LIHD). There were, however, marked differences among biomasses in terms of the effect of NaOH loading on xylose solubilization. Xylose solubilization from LIHD and CS increased almost linearly over all pretreatment levels — i.e. from 0 to 20 g NaOH per 100 g TS. For SG1, SG2, and BBS, values increased linearly between loadings of 5 and 20 g NaOH per 100 g TS. Interestingly, SG1, SG2, BBS, and MHW exhibited virtually no increases in xylose solubilizations when NaOH levels increased from 0 to 5 g NaOH per 100 g TS.

Solubilization of lignin increased with NaOH loading for all biomasses with maximum values occurring at the highest level of pretreatment [Fig 5.1(c)]. The highest alkali loading of 20 g NaOH per 100 g TS resulted in solubilizations of lignin that ranged from 30% (MHW) to 62% (CS). The relationship between alkali loading and lignin solubilization differed markedly among biomass types, but overall there appear to be two main behaviors. For LIHD and MHW, solubilization increased almost linearly up to an alkali loading of 10 g NaOH per 100 g TS, after which there was a leveling off. For the remaining biomasses (SG1, SG2, BBS, and CS), lignin solubilization increased almost linearly over all NaOH loadings. One explanation for these two behaviors is that LIHD contains woody biomasses in addition to grasses, legumes, and forbes (Sills and Gossett 2011; Tilman et al. 2006), and, thus, the lignins in LIHD and MHW may be chemically similar to each other and different from the lignins in the four grass species (SG1, SG2, BBS, and CS).

Values of TS solubilization increased almost linearly up to NaOH loadings of 10 g NaOH per 100 g TS for all biomasses [Fig. 5.1(d)]. After that, increases leveled off. Maximum values of TS solubilizations were observed at the highest level of pretreatment and they ranged from 17% (MHW) to 35% (LIHD). Additionally, for TS solubilizations, differences among biomasses were maintained across pretreatment levels.

5.3.2 Alkali Consumption

Alkali consumption versus alkali loading in units of g NaOH per 100 g TS is presented in Fig. 5.2. There are two reasons one might be interested in modeling alkali consumption. First, Pavlostathis and Gossett (1985) reported that alkali consumption determined pretreatment effectiveness for wheat straw, in subsequent anaerobic digestion. The second reason is of a

practical interest. At full scale, the low-temperature NaOH pretreatment described here is expected to include recycle of unconsumed alkali (Pavlostathis and Gossett 1985; Sills and Gossett 2011), and, thus, it would be advantageous to estimate alkali consumption *a priori*. NaOH consumption was also modeled with PLS regression, and the model is presented in Section 5.3.5.

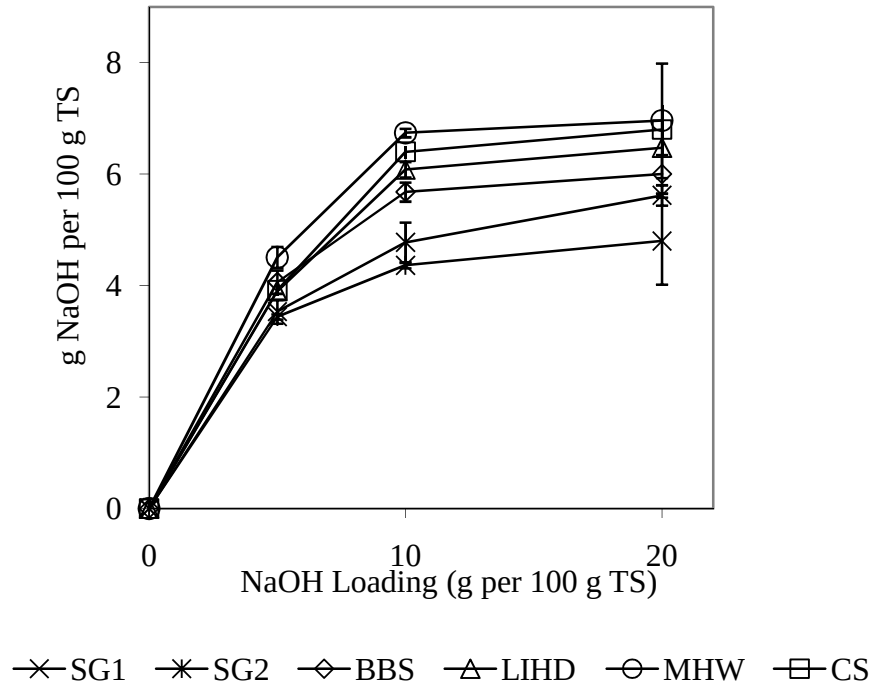

Figure 5.2 Alkali consumption versus NaOH loading for BBS, CS, LIHD, SG1, SG2, and MHW. Values are means of triplicate measurements and error bars represent standard deviations.

It can be observed in Fig. 5.2 that alkali consumption increased with NaOH loading, and maximum consumptions were observed at the highest pretreatment level for all six biomass types. This agrees with Pavlostathis and Gossett (1985) who measured alkali consumption over a wide range of alkali loadings during NaOH pretreatment of wheat straw. At the highest level of pretreatment, NaOH consumption ranged from 4.8 (SG2) to 7.0 (CS) g NaOH per 100 g TS. The relationship between alkali loading and alkali consumption was markedly similar for all biomass types— in all cases, consumption increased almost linearly up to a loading of 10 g NaOH per 100 g TS, and after that, values leveled off. Furthermore, as in the case of TS solubilization, differences among biomasses were maintained across pretreatment levels.

5.3.3 Effect of NaOH Loading on Overall Saccharification

Overall conversions of glucose and xylose through combined pretreatment and enzymatic hydrolysis for SG1, SG2, BBS, LIHD, MHW, and CS are presented in Fig. 5.3. Sugar conversions increased with NaOH loading for all biomasses with maximum conversions observed at the highest loading of 20 g NaOH per 100 g TS. At the highest pretreatment level, values of glucose conversions varied from 46% (MHW) to 82% (CS), and values of xylose conversions varied from 58% (MHW) to 75 and 77% (CS and LIHD, respectively). Glucose- and xylose-conversion data exhibited an S-curve relationship — as alkali loading increased from 0 to 5 g NaOH per 100 g TS, relatively small increases in sugar conversions were observed; as alkali loading increased from 5 to 10 g NaOH per 100 g TS, larger increases were observed; and as alkali loading increased to the highest pretreatment level of 20 g NaOH per 100 g TS, increases in sugar conversions lessened once again. Differences among biomass types were maintained among pretreatment levels, with one major exception — there was no increase in

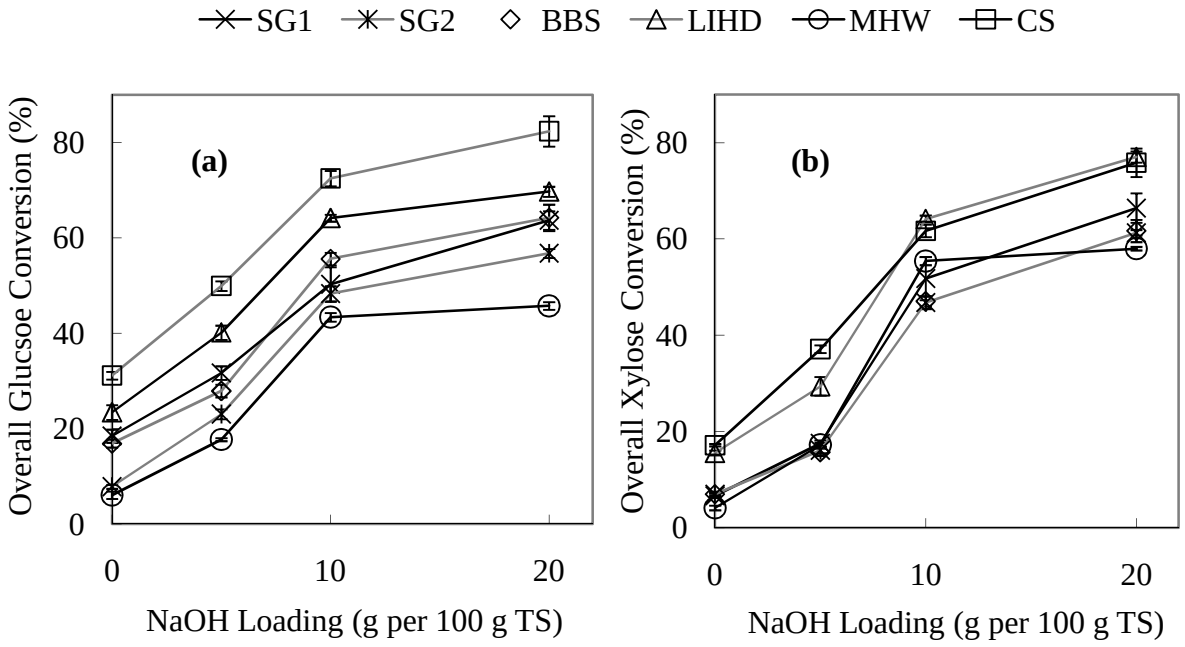

Figure 5.3 Effect of NaOH loading in pretreatment (20 ºC, 24 h) on overall glucose (a) and xylose (b) conversions from combined pretreatment and 72-h enzymatic hydrolysis of SG1, SG2, BBS, CS, LIHD, MHW, and CS. Values are averages of triplicate measurements and error bars represent standard deviations.

xylose conversion for MHW between pretreatment levels of 10 and 20 g NaOH per 100 g TS, while small increases were observed for the five other biomasses. At all pretreatment levels, minimum glucose and xylose conversions were associated with MHW, while maximum glucose conversions were associated with CS and maximum xylose conversions with CS and LIHD.

5.3.4 FTIR Spectra

FTIR-ATR spectra of raw biomass samples: SG1, SG2, BBS, LIHD, MHW, and CS are presented in Fig. 5.4. The spectra are fairly similar to one another, but MHW can be distinguished based on its relatively high absorbance at 1240 and 1740 cm^{-1}, and LIHD has a

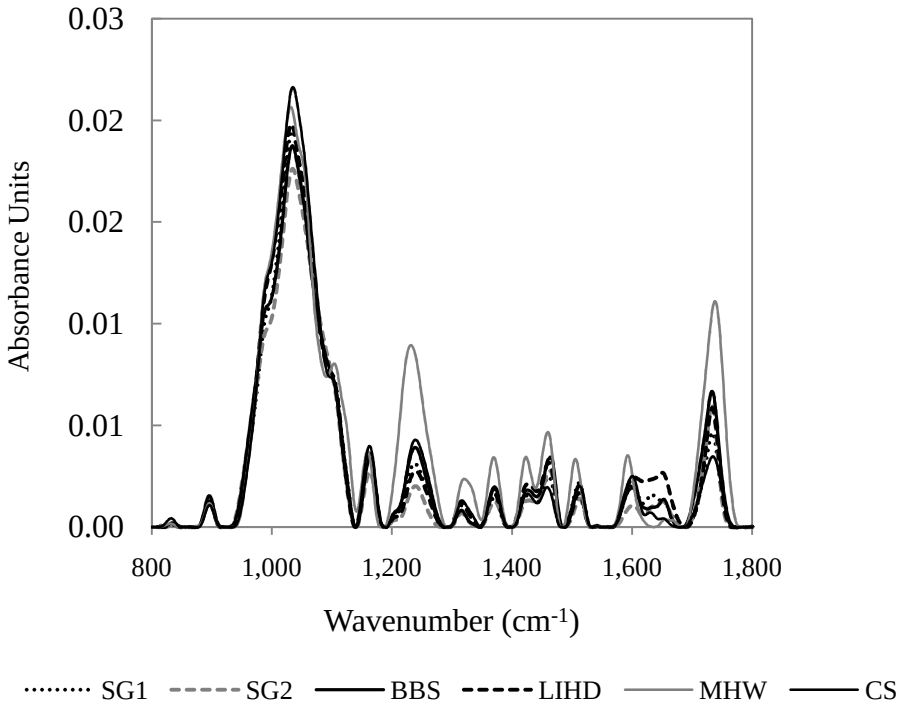

Figure 5.4 Average FTIR-ATR spectra (vector-normalized, ATR-corrected, and baseline-corrected) of untreated SG1, SG2, BBS, LIHD, MHW, and CS.

distinctive wide peak between approximately 1600 to 1700 cm^{-1}. The similarities of the six FTIR spectra illustrate the need for chemometric modeling for quantitative analysis.

5.3.5 PLS Regression Models — Pretreatment

PLS regression models were constructed with an X matrix consisting of the FTIR spectra of raw biomasses plus the level of NaOH loading (0, 5, 10, and 20 g per 100 g TS, scaled as described in Section 5.2.8.1) to predict solubilizations of glucose, xylose, lignin, and TS through pretreatment (Y vectors). An alkali consumption model was constructed with FTIR spectra of

raw biomasses plus three levels of pretreatment (5, 10, and 20 g NaOH per 100 g TS, logit-transformed as described in Section 5.2.8.1). We assessed the predictive ability of the models by considering values of Q^2 as well as differences between Q^2 and R^2. The value of R^2 provides an optimistic view of how well a model might predict for new samples, while Q^2 provides a pessimistic view (Wold et al. 2001). Furthermore, large differences between R^2 and Q^2 indicate over-fitting of multivariate regression models. It has been suggested that coefficient of determination values (in our case, Q^2) between 0.8 and 0.9 indicate models are of good predictive ability, and values higher than 0.9 indicate models are of excellent predictive ability (Tamaki and Mazza 2011). Additionally, Chen et al. (2010) suggested differences between R^2 and Q^2 of less than 0.2 indicate models of good predictive ability.

For glucose and xylose solubilizations through pretreatment, models with 5 LVs resulted in the lowest cross-validation errors (RMSECV) of 1.5% glucose and 2.3% xylose, and the highest Q^2 values of 0.86 for glucose and 0.90 for xylose. For lignin, TS, and alkali consumption, models with 3 LVs resulted in the lowest RMSECV values of 8.3% lignin, 3.2% TS, and 0.49 g NaOH per 100 g TS; and the highest Q^2 values of 0.79 for lignin, 0.85 for TS, and 0.83 for alkali consumption.

Plots of predicted versus measured values of glucose, xylose, lignin, and TS solubilized during pretreatment for LOO cross-validation and model calibration of PLS regression models with 5 LVs for glucose and xylose and 3 LVs for lignin and TS are presented in Figs. 5.5. and 5.6. Values of R^2 were 0.93 for glucose, 0.95 for xylose, 0.88 for lignin, and 0.91 for TS. Calibration errors (RMSEC) were 1.1% for glucose, 1.6% for xylose, 6.2% for lignin, and 2.4 % for TS. Similar results for the alkali-consumption model constructed with 3 LVs are presented in Fig. 5.7. RMSEC and R^2 values were 0.37 g NaOH per 100 g TS and 0.90, respectively.

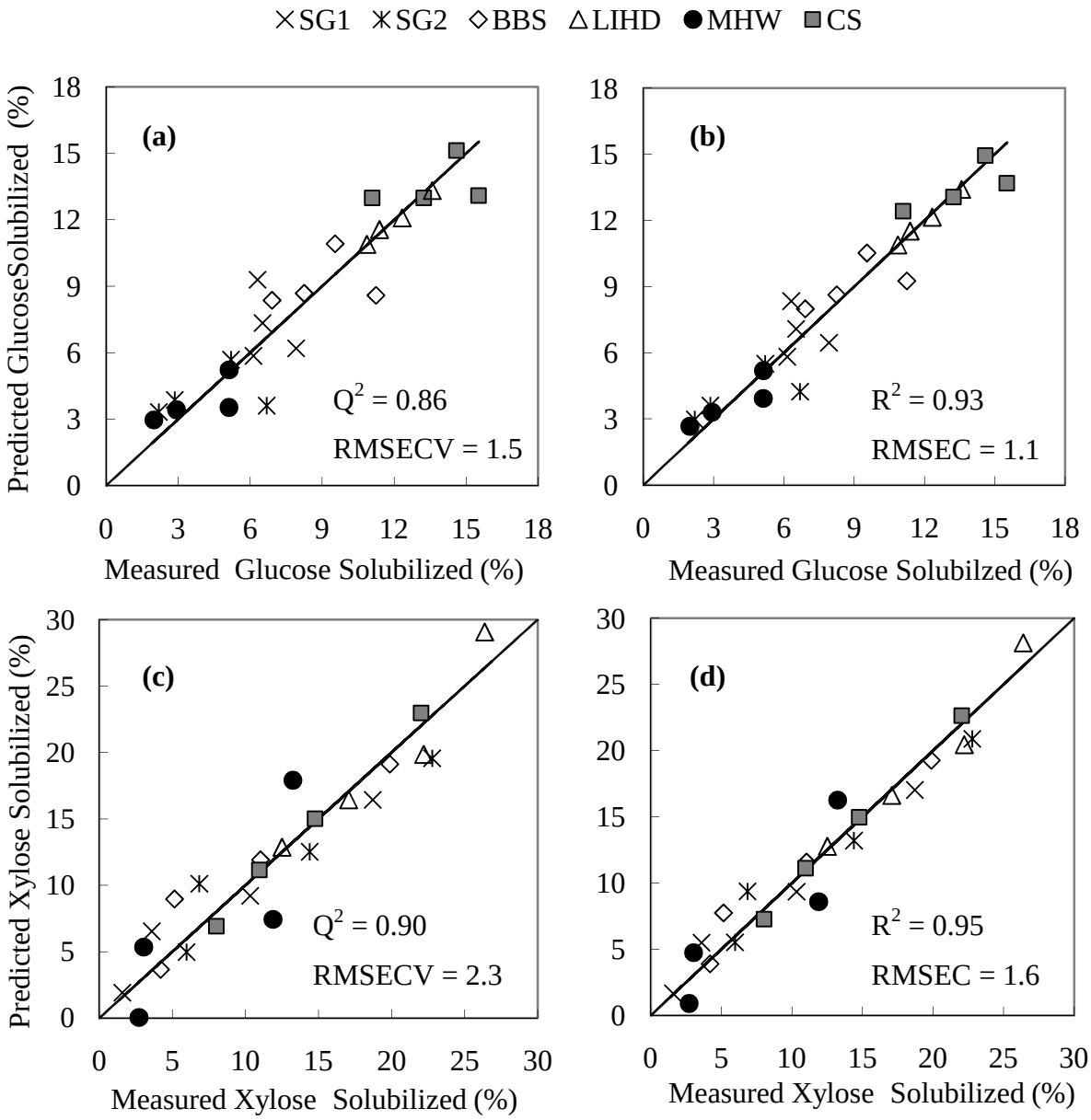

Figure 5.5 Predicted versus measured values of glucose (a and b) and xylose (c and d) solubilized in pretreatment for validation (a and c) and calibration (b and d) of PLS regression models constructed from spectra of raw SG1, SG2, BBS, LIHD, MHW, and CS and pretreatment levels of 0, 5, 10, and 20 g NaOH per 100 g TS. Glucose and xylose regression models were each constructed with 5 LVs.

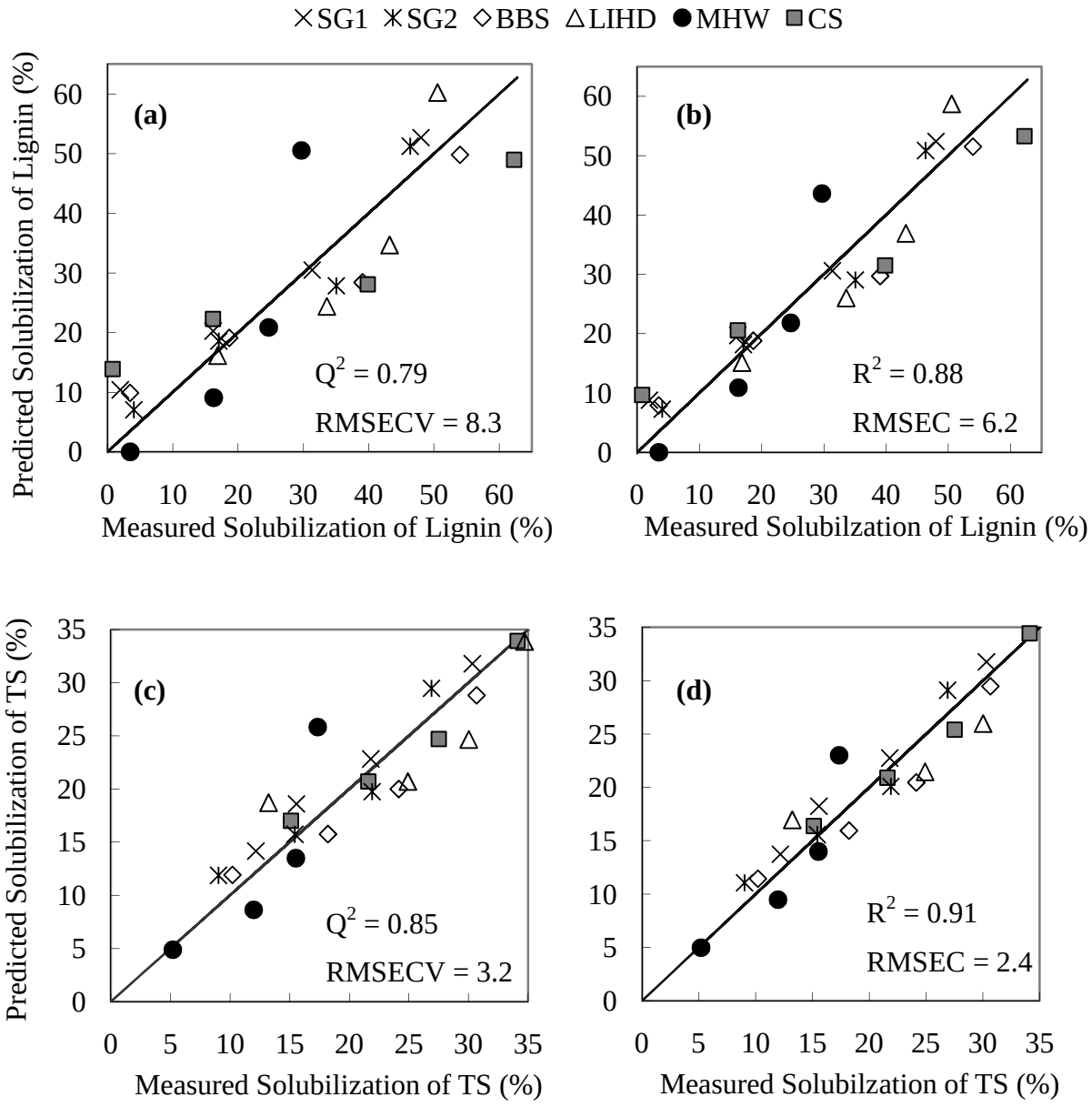

Figure 5.6 Predicted versus measured values of lignin (a and b) and TS (c and d) solubilized in pretreatment for validation (a and c) and calibration (b and d) of PLS regression models constructed from spectra of raw SG1, SG2, BBS, LIHD, MHW, and CS and pretreatment levels of 0, 5, 10, and 20 g NaOH per 100 g TS. Lignin and TS regression models were each constructed with 3 LVs.

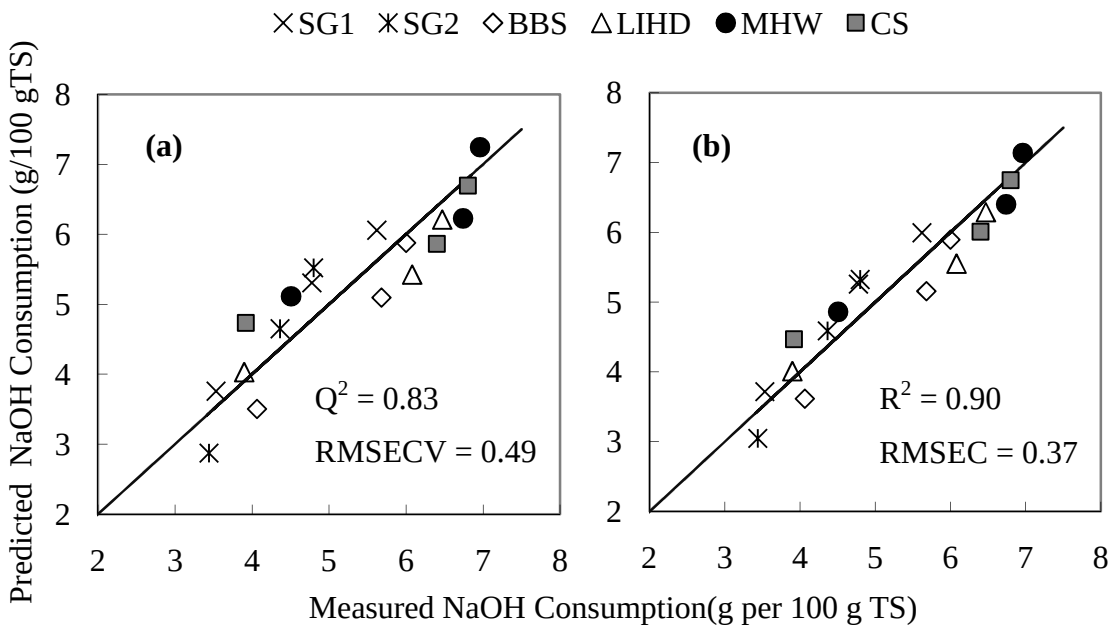

Figure 5.7 Predicted versus measured values of NaOH consumption for validation (a) and calibration (b) of PLS regression models constructed from spectra of raw SG1, SG2, BBS, LIHD, MHW, and CS and pretreatment levels of 5, 10, and 20 g NaOH per 100 g TS. The regression model was constructed with 3 LVs.

Glucose, xylose, TS, and alkali-consumption models are expected to predict new samples reasonably well — based on their values of Q^2, which are all between 0.8 and 0.9, and values of differences between R^2 and Q^2 of 0.07 for glucose, 0.05 for xylose, 0.06 for TS, and 0.07 for alkali-consumption (all lower than 0.2); however, the lignin model may not satisfactorily predict new samples, based on a Q^2 value of 0.79, which was slightly less than the criterion of 0.80 for good predictive ability.

We cannot provide a good explanation for the inferior fit for the lignin data, although this result is not surprising when one looks at the lignin solubilization data presented in Fig. 5.1(c). The data indicate that effects of pretreatment on lignin solubilization varied among biomasses to

a greater degree than for other cell-wall constituents. One would hope that the FTIR spectra of the raw biomasses could account for these differences and produce a satisfactory model. However, there may be physical characteristics of biomass samples (e.g. porosity) controlling solubilization of lignin that are not explained by FTIR spectra. An additional explanation for the relatively poor fit of the lignin model may be that lignin contained in hardwood is too different from lignin in the other biomasses. MHW pretreated with 20 g NaOH per 100 g TS fit the lignin model especially poorly [Figs. 5.6(a) and (b)], and Wold et al. 2001 reported that, in some cases, removing samples that differ significantly from the main population of samples significantly improves PLS-regression models. Still, considering that the models presented here were developed from only six spectra plus NaOH loading, these results are promising. It is likely that with additional biomass samples and perhaps pretreatment levels, PLS models that predict effects of NaOH pretreatment on all aspects of solids composition satisfactorily can be developed.

5.3.6 PLS Regression Models— Overall Sugar Conversions and Yields

PLS regression models were developed to predict overall glucose and xylose conversions and yields from combined pretreatment and enzymatic hydrolysis using FTIR spectra of untreated SG1, SG2, BBS, LIHD, MHW, and CS and three pretreatment levels: 5, 10, and 20 g NaOH per 100 g TS.

For predicting sugar conversions (% of potential glucose or xylose in raw biomasses), models with 3 LVs resulted in minimum RMSECV of 4.6 for glucose and 4.9 for xylose, and maximum Q^2 values of 0.93 for glucose and 0.94 for xylose. Validation and calibration plots for sugar conversions are presented in Fig. 5.8. Calibration results for conversion models resulted in

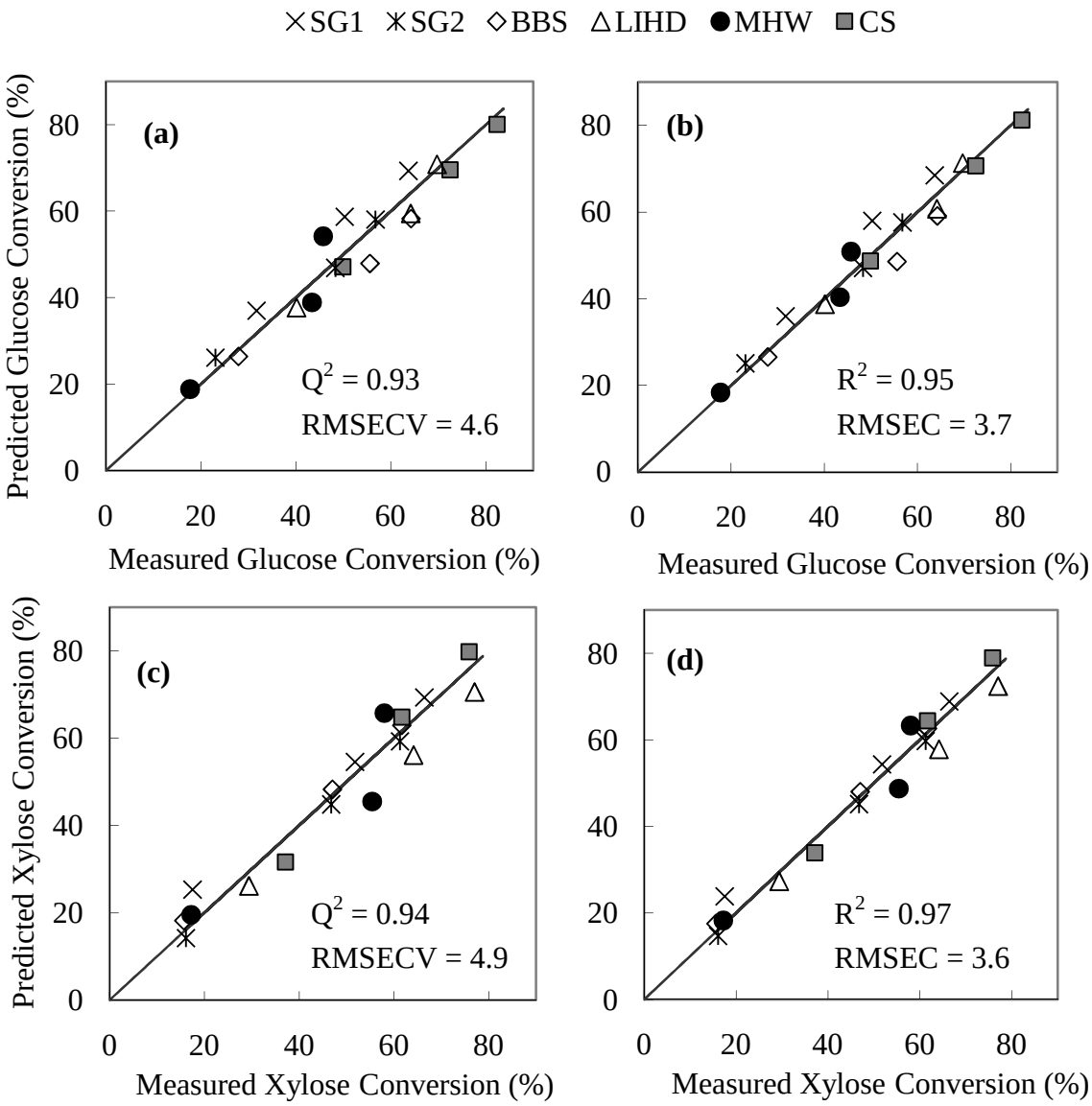

Figure 5.8 Predicted versus measured overall glucose (a and b) and xylose (c and d) conversions (from PT + hydrolysis) for validation (a and c) and calibration (b and d) of PLS regression models constructed from spectra of raw SG1, SG2, BBS, LIHD, MHW, and CS and pretreatment levels of 5, 10, and 20 g NaOH per 100 g TS. Glucose and xylose regression models were each constructed with 3 LVs.

RMSEC values of 3.7 for glucose and 3.6 for xylose and R^2 values of 0.95 for glucose and 0.97 for xylose. The high Q^2 values (above 0.90) indicated that sugar conversion models were of excellent predictive ability, and the low value of R^2 minus Q^2 — 0.02 for glucose and 0.03 for xylose — indicated that over-fitting was not a problem. A discussion of the regression coefficient matrices will be presented in a subsequent section.

Models with 3 LVs resulted in minimum values for the RMSECV and maximum values for Q^2 for sugar yields (glucose or xylose grams per 100 grams of raw TS)However, the fits presented in Fig. 5.9 were not quite as good compared with the conversion models (Fig. 5.8). Q^2 values were 0.88 for glucose and xylose yields (compared to 0.93 and 0.94 for glucose and xylose conversions). The inferior fits for the sugar-yield models are reasonable, since sugar-yield models must estimate the potential sugars in the raw substrates in addition to the extents of sugar conversion. Differences between R^2 and Q^2 were 0.05 for glucose yield and 0.06 for xylose yield — both smaller than 0.2. This combined with Q^2 values being between 0.8 and 0.9 indicated that glucose and xylose yields models are of good predictive ability.

5.3.8 Spectral Interpretation

For all PLS regression models, their respective dependent (Y) variables correlated significantly with NaOH level. In fact, VIP scores associated with pretreatment level (values ranged from 10 to 30) were one order of magnitude larger than VIP scores associated with spectral variables. However, VIP scores reflect importance in describing X as well as in describing Y. Therefore, high VIP scores alone do not ensure that pretreatment level is important for estimating Y. Clearly, pretreatment level is required for modeling X, since it is all that differentiates among each group of three or four samples associated with a specific biomass

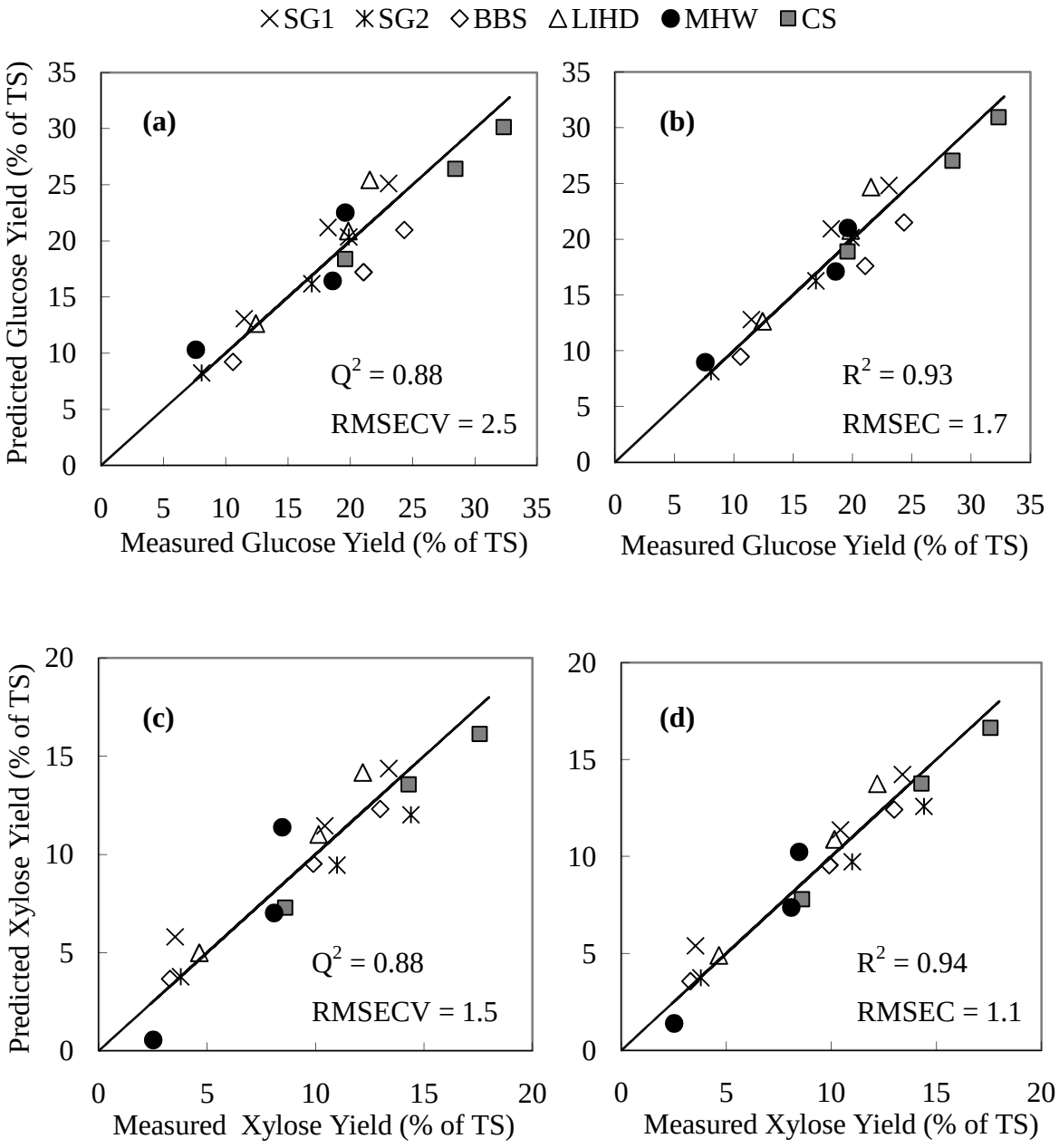

Figure 5.9 Predicted versus measured values of overall glucose (a and b) and xylose (c and d) yields (from PT + hydrolysis) for validation (a and c) and calibration (b and d) of PLS regression models constructed from spectra of raw SG1, SG2, BBS, LIHD, MHW, and CS and pretreatment levels of 5, 10, and 20 g NaOH per 100 g TS. Glucose and xylose regression models were each constructed with 3 LVs.

type. Additionally, all dependent variables (except, perhaps, glucose solubilized through pretreatment) were correlated with NaOH loading (Figs. 5.1, 5.2, and 5.3), and thus pretreatment level is unquestionably important for modeling Y.

What about the importance of FTIR spectra? Since pretreatment level and FTIR absorbances were of different scales, interpretation of their relative importance as predictive variables is not possible. However, since different biomass types clearly responded differently to NaOH pretreatment, then pretreatment level alone cannot be a good predictor. Biomass type matters, and FTIR is our presumed window into describing differences among biomasses. We can gain some insight into the relative importance of different parts of the IR spectrum by examining the regression coefficients of spectral variables alone.

Regression coefficients for spectral variables are presented in Figs. 5.10 through 5.14 for the models constructed and presented in this chapter. The 90% CIs of the coefficients and VIP scores are also depicted, as aids to interpretation. Scaling and variable transformation can alter both magnitude and 90% CI of regression coefficients. On the other hand, VIP scores are relatively unaffected by scaling or variable transformation. Thus, in spectral interpretation, we chose to use VIP scores to identify spectral regions of predictive importance. Since VIP scores are always positive, we then examined the 90% CIs of the coefficients in these VIP-score-identified, important spectral regions to determine whether the associated wavenumbers were associated positively or negatively with a model's particular dependent (Y) variable.

Previous authors suggest that VIP-score values > 1.0 indicate a variable is significant in the formulation of the PLS model (Andersen and Bro 2010; Chong and Jun 2005). We focused on variables with VIP-score values > 1.0, which in this study, were also associated with significant regression coefficients based on 90% CIs. For all PLS models, VIP scores > 1.0

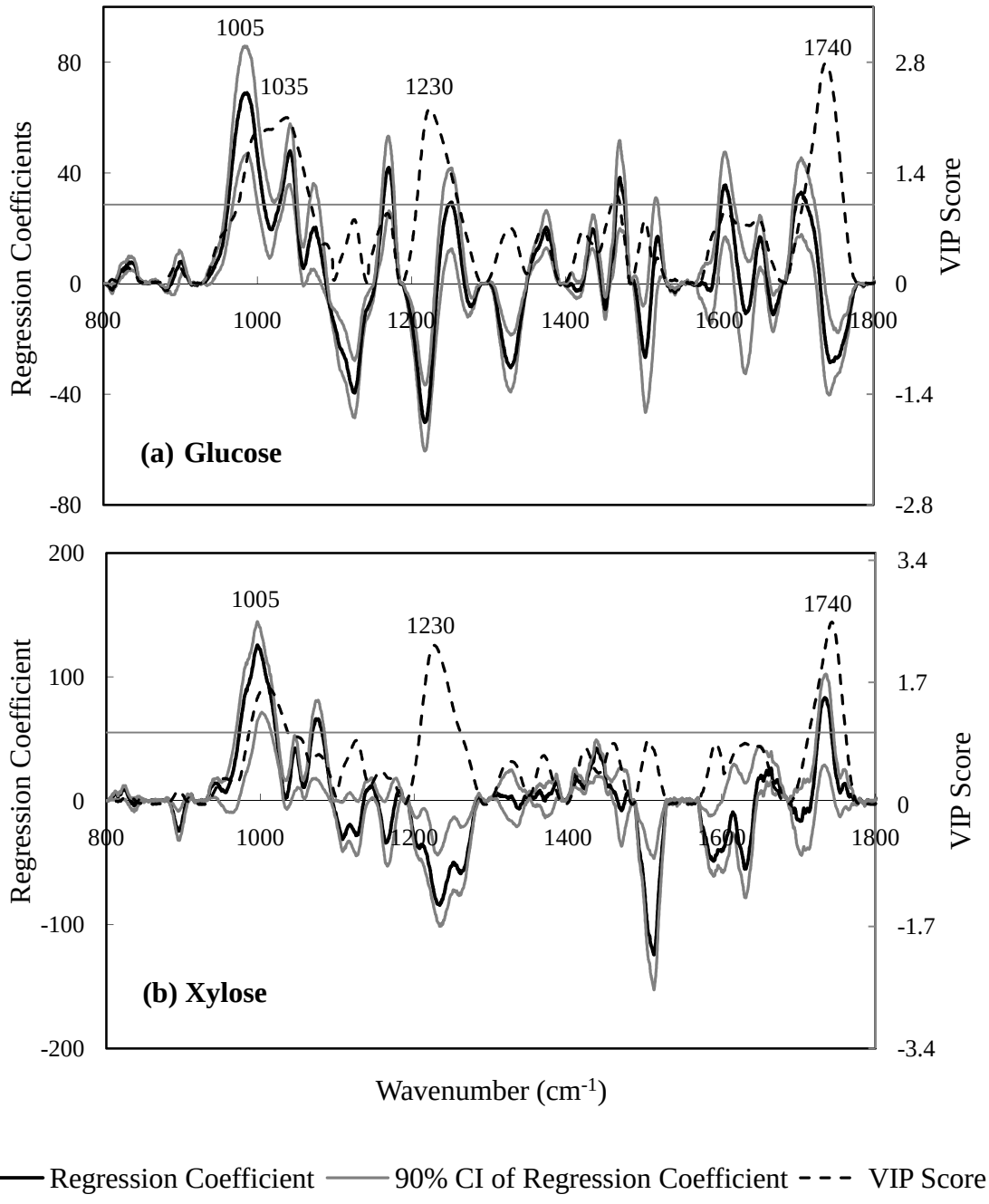

Figure 5.10 Regression-coefficient values and VIP scores versus wavenumber of PLS regression models for (a) glucose and (b) xylose solubilized through pretreatment. Wavenumber values with VIP-score values greater than 1.0 are listed above the associated peaks in units of cm^{-1}. Units of the y–axes are arbitrary.

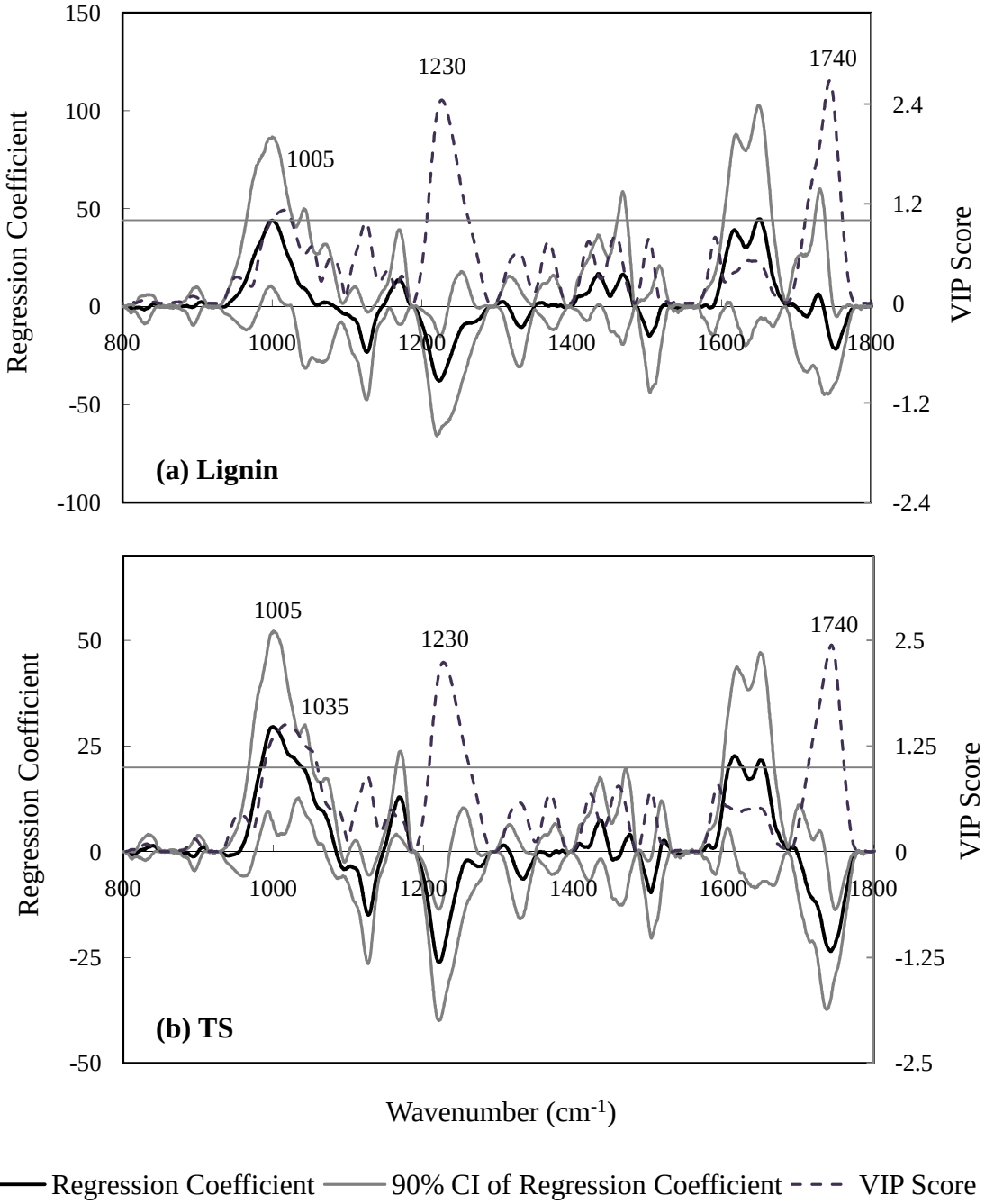

Figure 5.11 Regression-coefficient values and VIP scores versus wavenumber of PLS regression models for (a) lignin and (b) TS solubilized through pretreatment. Wavenumber values with VIP-score values greater than 1.0 are listed above the associated peaks in units of cm^{-1}. Units of the y–axes are arbitrary.

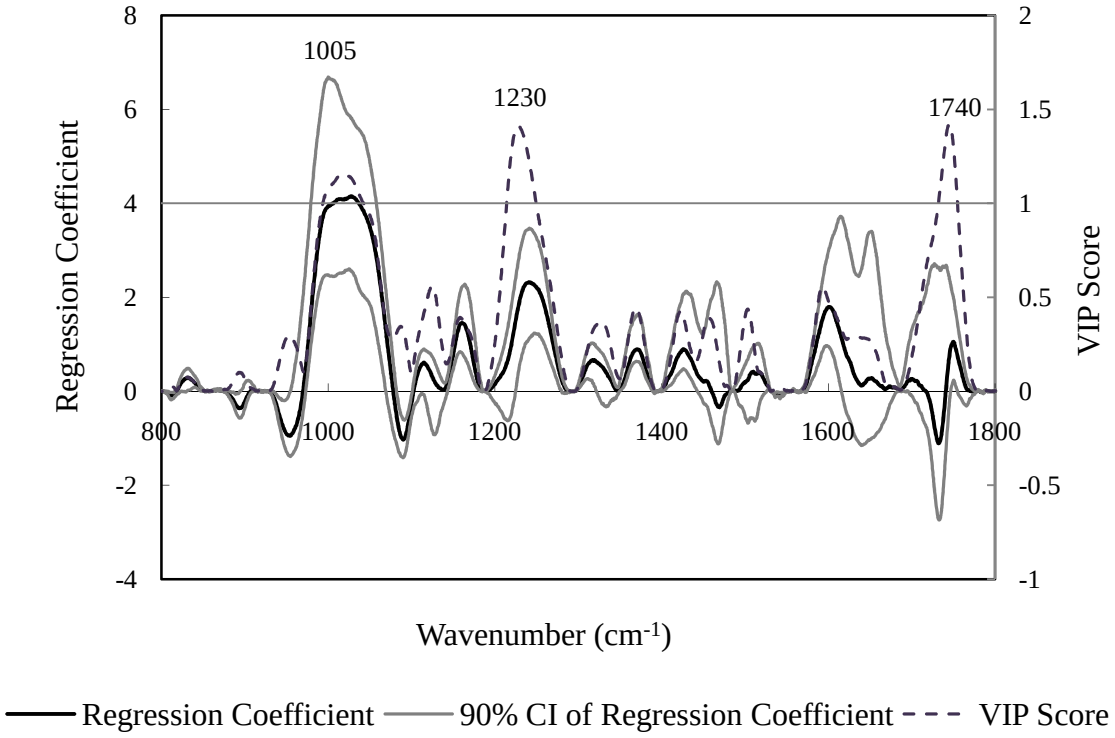

Figure 5.12 Regression-coefficient values and VIP scores versus wavenumber of PLS regression models for NaOH consumed during pretreatment. Wavenumber values with VIP-score values greater than 1.0 are listed above the associated peaks in units of cm^{-1}. Units of the y–axes are arbitrary.

came from the following group of four variables (though not all four were represented in all models): 1005, 1035, 1230, and 1740 cm^{-1}. Interpretation of the first two is elusive.

Regression coefficients associated with 1005 cm^{-1} had VIP > 1.0 for all models except the overall xylose conversion and the overall glucose and xylose yield models. The magnitudes of regression coefficients at 1005 cm^{-1} were positive in all instances. Absorbance in the region between 990 and 1010 has been correlated with a change in cellulose, from type I to type II, resulting from treatment with NaOH (Gwon et al. 2010; Nelson and O'Connor 1964). Given that our spectra were of raw solids, it is not clear why this wavenumber had significant VIP and

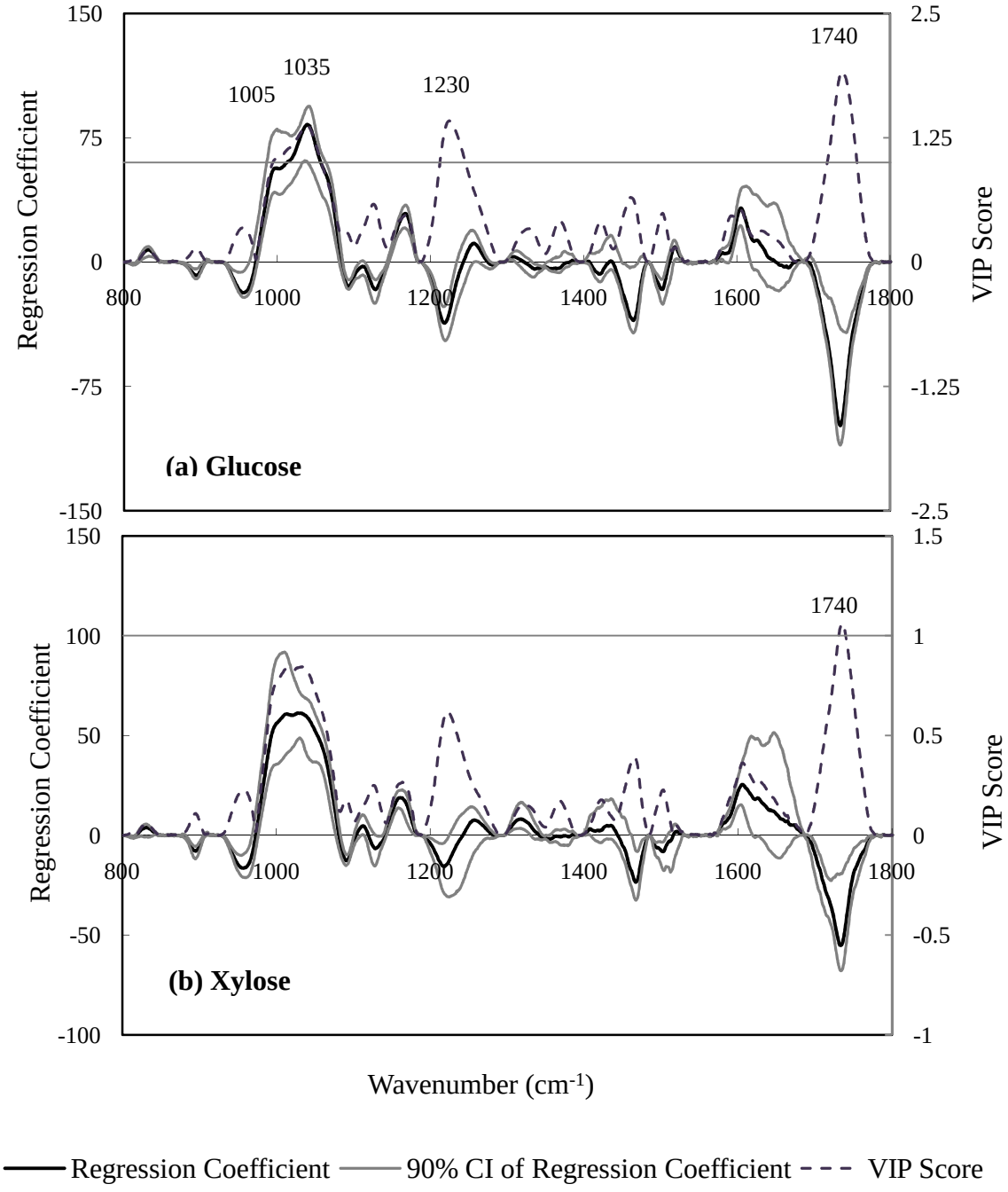

Figure 5.13 Regression-coefficient values and VIP scores versus wavenumber of PLS regression models for overall (a) glucose and (b) xylose conversions through pretreatment + enzymatic hydrolysis. Wavenumber values with VIP-score values greater than 1.0 are listed above the associated peaks in units of cm^{-1}. Units of the y–axes are arbitrary.

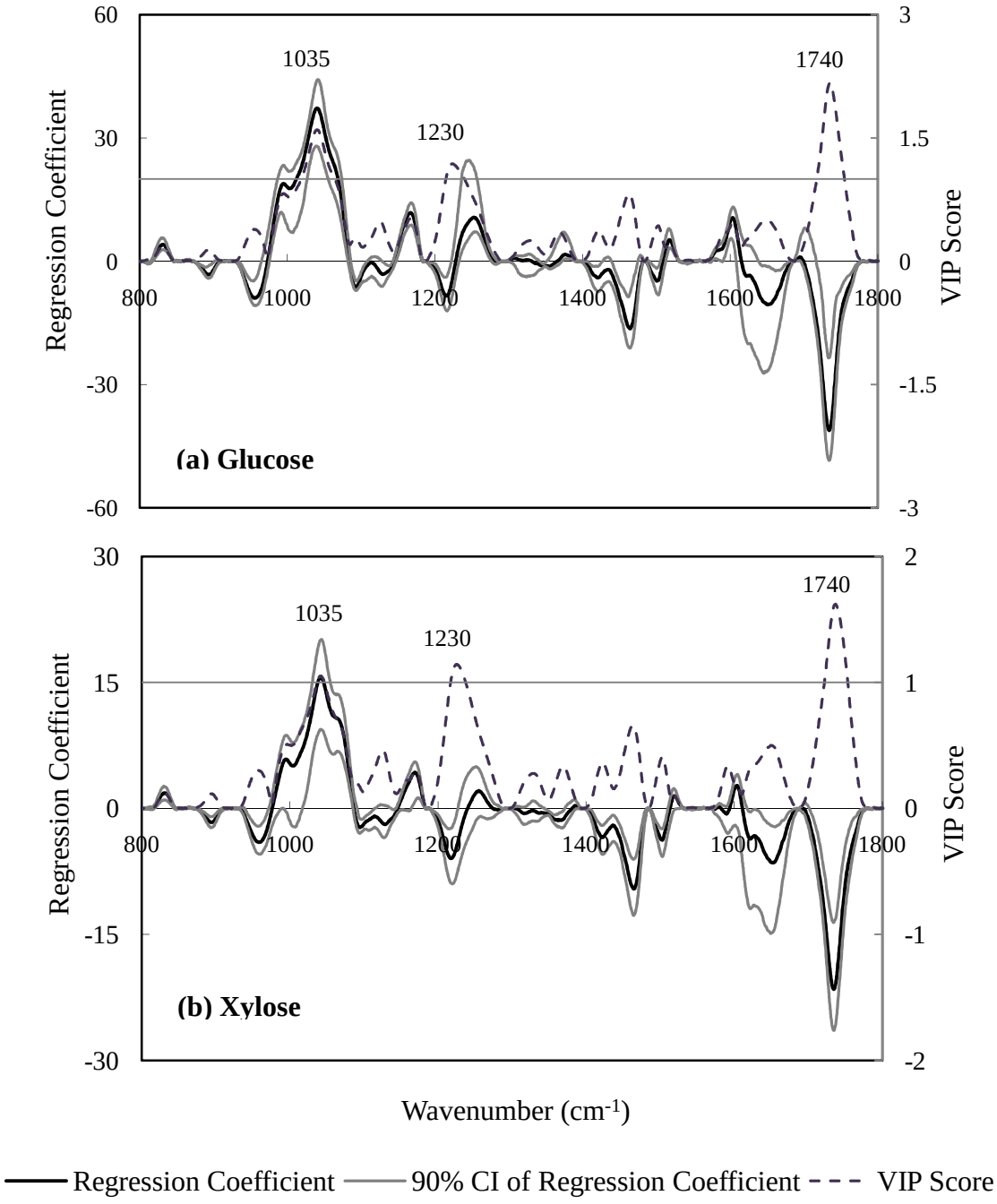

Figure 5.14 Regression-coefficient values and VIP scores versus wavenumber of PLS regression models for overall (a) glucose and (b) xylose yields (g sugar per 100 g TS or raw biomass) through pretreatment + enzymatic hydrolysis. Wavenumber values with VIP-score values greater than 1.0 are listed above the associated peaks in units of cm^{-1}. Units of the y–axes are arbitrary.

positive 90% CIs in most, but not all, of the models. peak lies in the general region associated with carbohydrates [900-1200 cm^{-1}] (McCann et al. 1997), making interpretation difficult.

Regression coefficients associated with 1035 cm^{-1} had VIP > 1.0 in models for solubilization of glucose and TS (but not xylose or lignin); for NaOH consumption; for glucose (but not xylose) conversion; and for glucose and xylose yield. In all instances, the magnitudes of the regression coefficients at 1035 cm^{-1} were positive. Absorbance at 1035 cm^{-1} has been attributed to pyranose ring stretching in cellulose (Brouchard et al. 1989), C-O and C-C bonds in hemicellulose (Wilson et al. 2000), and C-H, C-O, and C=O bonds in lignin (Faix 1992). Since it occurs in all lignocellulose component types, the reason for its importance in some (but not all) models is unclear.

Absorbance between 1220 and 1230 cm^{-1} has been attributed to C-C, C=O, and C-O bonds in condensed guaicyl lignin (Faix 1992), while absorbance at 1225 cm^{-1} has been attributed to O-H bending in crystalline cellulose (Nelson and O'Connor 1964). The magnitudes of the regression coefficients in this region were negative for all models except for alkali consumption. This makes sense— i.e. the presence of bonds that absorb in the region between 1220 and 1230 cm^{-1} limit depolymerization of lignocellulose; however, NaOH is consumed as a result of breaking bonds that limit saccharification.

The wavenumber 1740 cm^{-1}, which has been attributed to ester bonds (Owen and Thomas 1989), correlates negatively with glucose, lignin, and TS solubilizations in pretreatment, and overall glucose and xylose conversions and yields. Since ester bonds in lignocellulose link carbohydrate polymers to each other and to lignin (McCann and Carpita 2008), it seems reasonable that their presence in raw biomass would correlate negatively with cell-wall depolymerization in pretreatment and enzymatic hydrolysis. This wavenumber correlates

positively with xylose solubilization and alkali consumption in pretreatment. It is not clear why this wavenumber correlates positively with xylose solubilization in pretreatment. Perhaps, it is because the presence of ester bonds is attributed to the presence of hemicellulose, which in most biomasses is comprised predominantly of polymers with xylan backbones (Saha 2003). Since alkali consumption is due primarily to saponification reactions that break ester bonds (Pavlostathis and Gossett 1985), a positive correlation, in this case, is expected.

Even though there are difficulties in interpreting positive and negative magnitudes of spectral regression coefficients, the regression coefficients with values of VIP scores > 1.0 have been attributed to chemical features of lignocellulose. This suggests that the PLS models developed in this chapter are based on real chemistry and not chance correlation.

5.4 Conclusion

PLS regression was applied to FTIR spectra of six biomasses along with four levels of pretreatment (0, 5, 10, and 20 g NaOH per 100 g TS) to estimate the effect of pretreatment on solubilization of biomass components. Additionally, PLS regression models were constructed with FTIR spectra of the same six biomasses and three levels of pretreatment (5, 10, and 20 g NaOH per 100 g TS) to estimate NaOH consumption in pretreatment as well as overall glucose and xylose conversions (g sugars per 100 g potential sugars in raw biomasses) and overall glucose and xylose yields (g sugars per 100 g TS of raw biomasses). PLS models accurately predicted alkali consumption during pretreatment along with solubilizations of glucose, xylose, and TS. Moreover, we were able to construct models for overall sugar conversions with excellent predictive abilities and overall sugar yields with good predictive abilities.

These results are remarkable considering models were constructed primarily from six FTIR spectra. The ability to predict glucose and xylose yields without wet chemistry analysis suggests that this method could be very useful if works as well with other biomasses and pretreatments. Spectral attributions were tentative, but they indicate that the PLS models are based on chemical characteristics of lignocellulose that limit pretreatment and enzymatic hydrolysis. Further development of this technique may significantly reduce the need for wet-chemistry analyses and elucidate chemical features that contribute to plant cell wall recalcitrance.

Acknowledgements

This study was supported with funding from the Department of Transportation and the Northeast Sun Grant Initiative. We thank Larry Walker and Stephane Corgié at the Biofuels Research Laboratory, Cornell University, for providing analytical equipment and instruction in FTIR analysis; Deborah Ross and Michael Van Amburgh, Cornell University, for their facilities and instruction in biomass characterization; the Nanobiotchnology Center at Cornell University for use of the FTIR; Hillary Mayton from Cornell University, Matt McArdle from MESA Inc., James McMillan from NREL, Tom Stickle from Ligonier, PA, and Jason Hill and Matt Ziehr from the University of Minnesota for providing biomass samples; and Genencor Danisco for providing enzyme mixtures.

REFERENCES

Allison, GG, Thain, S.C., Morris, P. Morris, C., Hawkins, S., Hauck, B., Barraclough, T., Yates, N., Shield, I., Bridgwater, A., Donnison, I.S. 2009a. Quantification of hydroxycinnamic acids and lignin in perennial forage and energy grasses by Fourier-transform infrared spectroscopy and partial least squares regression. Bioresour Technol 100:1252-1261.

Allison GG, Thain SC, Morris P, Morris C, Hawkins S, Hauck B, Barraclough T, Yates N, Shield I, Bridgwater A and others. 2009b. Quantification of hydroxycinnamic acids and lignin in perennial forage and energy grasses by Fourier-transform infrared spectroscopy and partial least squares regression. Bioresour Technol 100:1252-1261.

Andersen CM, Bro R. 2010. Variable selection in regression— a tutorial. J Chemometrics 24:728-737.

Brouchard J, Abatzoglou N, Chornet E, Overend RP. 1989. Characterization of depolymerized cellulosic residues. Part I: Residues obtained by acid hydrolysis processes. Wood Sci Technol 23:343-355.

Chandra RP, Bura R, Mabee WE, Berlin A, Pan X, Saddler JN. 2007. Substrate pretreatment: The key to effective enzymatic hydrolysis of lignocellulosics? Adv Biochem Eng Biotechnol 108:67-93.

Chong IG, Jun CH. 2005. Performance of some variable selection methods when multicollinearity is present. Chemometrics Intelligent Lab Syst 78:103-112.

de Jong S. 1993. SIMPLS: an alternative approach to partial least squares regression. Chemometrics Intelligent Lab Syst 18:251-263.

Dodds DR, Gross RA. 2007. Chemistry: Chemicals from biomass. Science 318:1250-1251.

Efron B, Tibshirani R. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1:54-75.

Eriksson L, Johansson E, Kettaneh-Wold N, Wold S. 2001. Multi- and Mega Variate Data Analysis. Principles and Applications. Umea, Sweden: Umetrics AB.

Faix O. 1992. Fourier Transform Infrared Spectroscopy. In: Lin SY, Dence CW, editors. Methods in lignin chemistry. Berlin: Springer-Verlag. p 83-109.

Faust TD, Ibsen KN, Dayton DC, Hess JR, Kenney KE. 2008. The Biorefinery. In: Himmel ME, editor. Biomass recalcitrance. Deconstructing the plant cell wall for bioenergy. Oxford, United Kingdom: Blackwell Publisher Ltd. p 7-37.

Geladi P, Kowalski BR. 1986. Partial least-squares regression: a tutorial. Anal Chim Acta 185:1-17.

Gwon JG, Lee SY, Doh GH, Kim JH. 2010. Characterization of chemically modified wood fibers using FTIR spectroscopy for biocomposites. J Appl Polym Sci 116:3212-3219.

Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. 2007. Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science 315:804-807.

Keer DR, Meador JP. 1995. Modeling dose response using generalized linear models. Hazard/Risk Assessment 15:395-401.

Martens H, Naes T. 1989. Multivariate Calibration. Chichester: John Wiley & Sons Inc.

McCann MC, Carpita NC. 2008. Designing deconstruction of plant cell walls. Curr Opin Plant Biol 11:314-320.

McCann MC, Chen I, Roberts K, Kemsley EK, Sene C, Carpita NC, Stacey NJ, Wilson RH. 1997. Infrared microspectroscopy: Sampling heterogeneity in plant cell wall composition and architecture. Physiol Plant 100:729-738.

Merino ST, Cherry J. 2007. Progress and challenges in enzyme development for biomass utilization. Adv Biochem Eng Biotechnol 108:95-120.

Mevik BH, Wehrens R. 2007. The pls package: principal component and partial least squares regression in R. J Stat Softw 18:1-24.

Mooney CA, Mansfield SD, Touhy MG, Saddler JN. 1998. The effect of initial pore volume and lignin content on the enzymatic hydrolysis of softwoods. Bioresour Technol 64:113-119.

Nelson M, O'Connor RT. 1964. Relation of certain infrared bands to cellulose crystallinity and crystal lattice type. Part I. Spectra of lattice types I, II, III, and of amorphous cellulose. J Appl Polym Sci 8:1311-1324.

Ohgren K, Bura R, Saddler J, Zacchi G. 2007. Effect of hemicellulose and lignin removal on enzymatic hydrolysis of steam pretreated corn stover. Bioresour Technol 98:2503-2510.

Owen NL, Thomas DW. 1989. Infrared studies of "hard" and "soft" woods. Appl Spectrosc 43:451-455.

Pauly M, Keegstra K. 2008. Cell-wall carbohydrates and their modification as a resource for biofuels Plant J 54:559-568.

Pauly M, Keegstra K. 2010. Plant cell wall polymers as precursors for biofuels. Curr Opin Plant Biol 13:304-311.

Pavlostathis SG, Gossett JM. 1985. Alkaline treatment of wheat straw for increasing anaerobic biodegradability. Biotechnol Bioeng 27:334-344.

Raiskila S, Pulkkinen M, Laakso T, Fagerstedt K, Löija M, Mahlberg R, Paajanen L, Ritschkoff A-C, Saranpää P. 2007. FTIR spectroscopic prediction of Klason and acid soluble lignin variation in Norway spruce cutting clones. Silva Fennica 41:351-371.

Saha BC. 2003. Hemicellulose bioconversion. J Ind Microbiol Biotechnol 30:279-291.

Sills DL, Gossett JM. 2011. Assessment of commercial hemicellulases for saccharification of alkaline pretreated perennial biomass. Bioresour Technol 102:1389-1398.

Somerville C. 2006. The billion-ton biofuels vision. Science 312:1277.

Stone M. 1974. Cross validatory choice and assessment of statistical predictions. J Stat Soc B 36:111-147.

Tamaki Y, Mazza G. 2011a. Rapid determination of carbohydrates, ash, and extractives contents of straw using attenuated total reflectance Fourier transform mid-infrared spectroscopy. J Agric Food Chem 59:6346-6352.

Tamaki Y, Mazza G. 2011b. Rapid determination of lignin content of straw using attenuated total reflectance Fourier transform mid-infrared spectroscopy. Journal of Agricultural and Food Chemistry 59:504-512.

Tilman D, Hill J, Lehman C. 2006. Carbon negative biofuels from low-input high-diversity grassland biomass. Science 314:1598-1600.

Wehrens R. 2011. Chemometrics with R. Multivariate data analysis in the natural sciences and life sciences. Gentleman R, Hornik K, Parmigiani G, editors. Berlin: Springer-Verlag Berlin and Heidelberg.

Wilson RH, Smith A, Kacuráková M, Saunders K, Wellner N, Waldron KW. 2000. The mechanical properties and molecular dynamics of plant cell wall polysaccharides studied by Fourier-transform infrared spectroscopy. Plant Phys 124:397-405.

Wold S, Sjostrom M, Eriksson L. 2001. PLS-regression: a basic tool of chemometrics. Chemometrics Intelligent Lab Syst 58(2):109-130.

Yang B, Wyman CE. 2004. Effect of xylan and lignin removal by batch and flowthrough pretreatment on the enzymatic digestibility of corn stover cellulose. Biotechnol Bioeng 86:88-98.

Zhao Y, Wang Y, Zhu JY, Ragauskas A, Deng Y. 2008. Enhanced Enzymatic Hydrolysis of Spruce by Alkaline Pretreatment at Low Temperature. Biotechnol Bioeng 99:1320-1328.

Zhou G, Taylor G, Polle A. 2011. FTIR-ATR-based prediction and modeling of lignin and energy contents reveals independent intra-specific variation of these traits in bioenergy poplars. Plant Methods 7(9) Available at http://www.plantmethods.com/content/7/1/9.
.

CHAPTER 6

CONCLUSIONS

6.1 Summary of Research

Two major objectives are addressed in this dissertation. The first objective was to ascertain whether hydrolysis must be customized to different biomasses that have the potential to be major feedstocks for biorefineries. This was addressed through studies, described in Chapter 3, conducted with two biomasses — switchgrass and a low-impact, high-diversity mixture of prairie biomasses (LIHD). The second major objective was to ascertain the potential for using Fourier transform infrared spectroscopy (FTIR), combined with partial least squares (PLS) regression modeling, to replace wet-chemistry analyses and enzymatic assays in estimating saccharification from various biomasses. Two studies, described in Chapters 4 and 5, demonstrated that FTIR with PLS modeling could usefully predict sugar productions from pretreatment and enzymatic hydrolysis.

6.1.1 Must Hemicellulase Supplementation be Customized for NaOH Pretreated Switchgrass and LIHD?

In hydrolysis of NaOH-pretreated switchgrass and LIHD, hemicellulase supplementation was required to achieve maximum sugar yields. Furthermore, the data support a model that considers the actions of commercial cellulase and commercial hemicellulase mixtures as contributing independently to saccharification of NaOH-pretreated switchgrass and LIHD -- except at low cellulase enzyme loadings, where synergistic effects were noted. The data and model indicate that, although absolute sugar yields differed between the two substrates, their *marginal enzyme effectiveness* (incremental sugar produced per incremental enzyme added) was

the same over a wide range of CB-loadings. This suggests that it is not necessary to customize hemicellulase supplementation and cellulase loadings for the two substrates studied here. Whether this finding is true of other biomasses and other pretreatment technologies remains to be investigated.

Indeed, the most significant practical contribution of this portion of study was this novel way of looking at the enzymatic process — the concept of marginal effectiveness. Too many of the saccharification studies in the literature seem to have had, as their objective, maximizing sugar yields regardless of the cost. The marginal effectiveness plot, presented in Fig. 4.5, clearly demonstrates that it is not cost effective to add unreasonably high enzyme loadings required to release all of the sugar contained in pretreated plant substrates. A full-scale biorefinery will need to minimize cost, not maximize sugar production.

6.1.2 FTIR Spectra of NaOH Pretreated Biomass Accurately Predict Saccharification in Enzymatic Hydrolysis.

FTIR-ATR spectroscopy combined with PLS regression was used to estimate 72-h glucose and xylose conversions and yields from enzymatic hydrolysis of six types of lignocellulosic biomass that underwent four levels of alkaline pretreatment — a data set providing a broad range of conversions and yields. We were able to construct PLS-regression models that accurately predicted glucose and xylose conversions (g sugars per 100 g potential sugars), as well as glucose and xylose yields (g sugars per 100 g TS). The ability to predict sugar yields from FTIR spectra of pretreated solids alone, without wet chemical analyses, suggests that this method could be very useful if the approach works equally well when applied to other biomasses and other pretreatment technologies.

Regions of the IR spectrum that were retained after variable selection could, in all instances, be identified with known chemical functional groups in lignocellulose. For example, three wavenumbers attributed to lignin — 1465, 1500, and 1595 cm^{-1} (Faix 1992; Owen and Thomas 1989) — correlated negatively with glucose and xylose conversions and yields, which is expected, since lignin is known to limit saccharification of pretreated lignocellulose. Additionally, absorbance in the region of 1720 to 1760 cm^{-1}, which have been attributed to ester groups (Faix 1992; Owen and Thomas 1989), correlated negatively (1730 cm^{-1}) and positively (1750 cm^{-1}) with glucose and xylose productions. This result also makes sense, since absorbance at 1730^{-1} has been attributed to intact ester bonds (Owen and Thomas 1989), known to limit saccharification (Grabber 2005; Ohgren et al. 2007), and absorbance at 1755 cm^{-1} has been attributed to free ester groups (Owen and Thomas 1989), suggesting ester-link breakage, previously correlated with increased digestibility of alkaline pretreated wheat straw (Pavlostathis and Gossett 1985). Wavenumbers between 800 and 1440 cm^{-1} are attributed to bonds present in multiple plant-cell-wall polymers making this area of the spectrum extremely difficult to interpret (Faix 1992; Hulleman et al. 1994). Spectral attributions presented in Table 4.2, while tentative, indicate that the PLS models are based in chemical reality, and that FTIR spectra contain chemical information that correlates with recalcitrance.

6.1.3 FTIR Spectra of Raw Biomasses Accurately Predicted Effects of NaOH Pretreatment

PLS regression modeling was applied to FTIR spectra of six raw biomasses along with four levels of pretreatment (0, 5, 10, and 20 g NaOH per 100 g TS) to estimate the effect of pretreatment on solubilizations of biomass components. Additionally, PLS regression models were constructed with FTIR spectra of the same six biomasses and three levels of pretreatment

(5, 10, and 20 g NaOH per 100 g TS) to estimate NaOH consumption in pretreatment as well as overall glucose and xylose conversions (g sugars per 100 g potential sugars in raw biomasses) and overall glucose and xylose yields (g sugars per 100 g TS of raw biomasses) from combined pretreatment and enzymatic hydrolysis. PLS models accurately predicted alkali consumption during pretreatment along with solubilizations of glucose, xylose, and TS. Moreover, we were able to construct models for overall sugar conversions with excellent predictive abilities and overall sugar yields with good predictive abilities.

These results are remarkable considering models were constructed primarily from six FTIR spectra. The ability to predict glucose and xylose yields without wet chemistry analysis suggests that this method could be very useful if works as well with other biomasses and pretreatments. The results of this study suggest that the "chemical fingerprint" created through FTIR spectroscopy provides all of the information needed to predict saccharification from combined pretreatment and enzymatic hydrolysis. Whatever physical characteristics might also be envisioned as important to prediction appear to have been adequately captured in the infrared spectra.

The FTIR studies presented here suggest that it is possible to use FTIR to customize pretreatment and enzymatic hydrolysis. FTIR spectra of raw biomasses, which can be produced in a few minutes, might allow biorefineries to choose conditions in pretreatment and enzymatic hydrolysis needed for optimal sugar productions.

6.2 Suggestions for Future Research

The study described in Chapter 3, which determined that hemicellulase mixtures do not need to be customized to NaOH pretreated switchgrass and LIHD should be extended to include additional biomasses and pretreatments. Moreover, this study was limited by the use of hemicellulase mixtures with multiple activities. Therefore, it would be advantageous to conduct similar studies that measure the effect of supplementing cellulase mixtures with pure hemicellulases with known activities. This may help determine if specific hemicellulase activities are required across biomass types, or if biomasses vary in their requirements for hemicellulase activities. Additionally, studies with pure hemicellulases may elucidate mechanisms by which pretreated lignocellulose resists biological depolymerization.

The sugar yield models developed from spectra of raw biomasses, developed in Chapter 5, should be extended to include additional biomasses, pretreatments, as well as enzyme mixtures and loadings. Since this model was able to predict sugar yields (g sugar per g TS) without *a-priori* knowledge of biomass composition, it may prove to be very useful for the practitioner. In fact, it may be possible to develop models that will estimate sugar yields from FTIR spectra of raw biomasses given different conditions in pretreatment and enzymatic hydrolysis. This could aid in the design of more effective saccharification processes, and possibly make it easier to tailor pretreatment and hydrolysis conditions on multiple feedstocks in one bioprocessing facility.

REFERENCES

Faix O. 1992. Fourier Transform Infrared Spectroscopy. In: Lin SY, Dence CW, editors. Methods in lignin chemistry. Berlin: Springer-Verlag. p 83-109.

Grabber JH. 2005. How do lignin composition, structure, and cross-linking affect degradability? A review of cell wall model studies. Crop Sci 45:820-831.

Hulleman SHD, van Hazendonk JM, van Dam JEG. 1994. Determination of crystallinity in native cellulose from higher plants with diffuse reflectance Fourier infrared spectroscopy. Carbohydr Res 261:163-172.

Ohgren K, Bura R, Saddler J, Zacchi G. 2007. Effect of hemicellulose and lignin removal on enzymatic hydrolysis of steam pretreated corn stover. Bioresour Technol 98:2503-2510.

Owen NL, Thomas DW. 1989. Infrared studies of "hard" and "soft" woods. Appl Spectrosc 43:451-455.

Pavlostathis SG, Gossett JM. 1985. Alkaline treatment of wheat straw for increasing anaerobic biodegradability. Biotechnol Bioeng 27:334-344.

APPENDIX 1

BIOMASS SPECIES IN LIHD

Table A.1 The 38 species planted in the plot from which LIHD came.

Species	Functional Type
*Achillea millefolium**	Forb
Agropyron repens	C_3 grass
Agropyron smithii	C_3 grass
*Andropogon gerardi**	C_4 grass
*Amorpha canescens**	Woody legume
*Asclepias tuberosa**	Forb
Aster azureus	Forb
Astragalus Canadensis	Legume
*Baptisia leucantha**	Legume
*Bouteloua curtipendula**	C_4 grass
*Bouteloua gracilis**	C_4 grass
Bromus inermis	C_3 grass
Buchloe dactyloides	C_4 grass
Calamagrostis canadensis	C_3 grass
Coreopsis palmata	Forb
Elymus canadensis	C_3 grass
Euphorbia corollata	Forb
*Koeleria cristata**	C_3 grass
Leersia oryzoides	C_3 grass
*Lespedeza capitata**	Legume
*Liatris aspera**	Forb
*Lupinus perennis**	Legume
*Monarda fistulosa**	Forb
*Panicum virgatum**	C_4 grass
*Petalostemum candidum**	Legume
*Petalostemum purpureum**	Legume
*Petalostemum villosum**	Legume
*Poa pratensis**	C_3 grass
Rudbeckia hirta	Forb
*Schizachyrium scoparium**	C_4 grass
*Solidago nemoralis**	Forb
*Solidago rigida**	Forb
*Sorghastrum nutans**	C_4 grass
*Sporobolus cryptandrus**	C_4 grass
Stipa comata	C_3 grass
Stipa spartea	C_3 grass
Vicia villosa	Legume
Zizia aurea	Forb

*Species is likely present in LIHD mix used in this study.

APPENDIX 2

CELL WALL CONSTITUENTS SOLUBILIZED THROUGH ALKALINE PRETREATMENT

Table A.2 Total solids (TS), glucan, xylan, and lignin solubilized through pretreatment (mean value of triplicates ± standard deviations), % of original constituent.

		Switchgrass 1	Switchgrass 2	Big Blue Stem	LIHD	Mixed Hardwood	Corn Stover
0 g NaOH /100 g TS	TS	12.2 ± 0.1	9.0 ± 1.1	10.2 ± 0.4	13.2 ± 0.5	5.20 ± 0.1	15.1 ± 0.2
	Glucan	6.14 ± 0.5	2.20 ± 0.4	6.92 ± 0.7	10.9 ± 1.1	1.99 ± 0.7	11.1 ± 0.7
	Xylan	1.58 ± 0.4	5.97 ± 0.3	4.21 ± 0.5	12.5 ± 0.5	2.71 ± 0.5	8.02 ± 0.3
	Lignin	1.96 ± 0.4	4.02 ± 1.72	3.45 ± 0.7	16.9 ± 0.4	3.48 ± 0.3	0.80 ± 0.8
5 g NaOH /100 g TS	TS	15.6 ± 0.0	15.4 ± 0.4	18.2 ± 0.3	24.9 ± 1.1	12.0 ± 0.2	21.6 ± 0.3
	Glucan	7.92 ± 0.5	2.86 ± 0.3	8.25 ± 1.2	11.4 ± 0.3	2.93 ± 0.2	13.2 ± 0.8
	Xylan	3.60 ± 0.4	6.85 ± 0.3	5.15 ± 0.7	17.1 ± 1.9	3.04 ± 0.2	11.0 ± 0.4
	Lignin	16.2 ± 0.7	17.1 ± 0.4	18.6 ± 0.4	33.6 ± 0.4	16.3 ± 0.5	16.2 ± 0.7
10 g NaOH /100 g TS	TS	21.8 ± 0.8	21.9 ± 0.1	24.1 ± 0.2	30.0 ± 0.8	15.5 ± 0.3	27.5 ± 0.3
	Glucan	6.52 ± 0.5	6.69 ± 0.8	11.2 ± 0.4	12.3 ± 0.3	5.12 ± 0.7	15.5 ± 0.8
	Xylan	10.3 ± 0.6	14.4 ± 0.4	11.0 ± 0.2	22.2 ± 0.5	11.9 ± 0.3	14.8 ± 0.3
	Lignin	31.4 ± 1.2	35.1 ± 1.3	39.1 ± 0.6	43.2 ± 0.3	24.7 ± 0.5	29.7 ± 0.4
20 g NaOH /100 g TS	TS	30.3 ± 1.2	26.9 ± 1.5	30.7 ± 0.7	34.7 ± 1.7	17.4 ± 0.1	34.1 ± 0.6
	Glucan	6.31 ± 1.3	5.20 ± 0.3	9.54 ± 1.2	13.6 ± 0.6	5.12 ± 0.6	14.6 ± 1.1
	Xylan	18.7 ± 1.1	22.8 ± 0.2	19.9 ± 0.6	26.4 ± 1.0	13.2 ± 0.2	22.0 ± 0.3
	Lignin	48.0 ± 2.0	46.4 ± 0.5	54.0 ± 0.4	50.6 ± 0.7	29.7 ± 0.4	62.3 ± 0.7